nature

www.nature.com/nature

Vol 452 | Issue no. 7190 | 24 April 2008

# New sources of sex cells

Within the next decade or so, it will become possible to derive sperm and eggs from skin cells. The ethical and technical hurdles need to be addressed with the scientific and therapeutic benefits in mind.

Earlier this month, the world had its first look at a pregnant man, a jarring reminder of how conventions in the way humans are created could shift. Just a week later, scientists, bioethicists, lawyers and journal editors convened in Hinxton, UK, to ponder how long it will be until sperm and eggs can be made entirely in a Petri dish from, say, skin cells induced to pluripotency. They asked questions such as how will these advances transform reproductive research and medicine? How might they change society? Is the bioethics community prepared?

These questions have acquired a sense of urgency, at least in Britain, where a government bill updating the 1990 Human Fertilisation and Embryology Act is working its way through the system. If passed by the House of Commons in its present form, the bill would allow basic research on human gametes derived *in vitro* but would not permit the use of such gametes for fertility treatment. Opponents may yet propose amendments designed to stop all embryo research — including the demonstration that *in-vitro*-derived gametes are normal by seeing whether they are capable of fertilization and further development. The Hinxton group's consensus statement, arrived at after three days of discussion, can be viewed at www.hinxtongroup.org/Consensus_HG08_FINAL.pdf.

The organizers scheduled the topic of pluripotent stem cell derived gametes (PSCDGs) two years ago. Since then, making induced pluripotent stem (iPS) cells has become relatively straightforward, but although the technique circumvents the ethical and technical problems associated with collecting human eggs and creating embryos, there are still serious ethical implications. Indeed, the facility with which iPS cells can be derived could make it easier to derive gametes from any person, living or dead.

The potential benefits of PSCDGs are impressive. They could reveal much about the mechanism of gamete development, with implications for treating infertility, certain cancers and genetic diseases, and for facilitating drug tests on gametes. Other potential applications present social and ethical challenges: germline genetic modification for the correction of disease mutations or for research; various forms of biological enhancement; genetic screening of embryos for selection; and enabling same-sex couples to have their own genetic offspring.

The Hinxton group believes that eggs and sperm will be generated from pluripotent stem cells in 5–15 years, although it concedes this is something of a guess. Technically, a major sticking point is getting the developing gametes to undergo meiosis. Ethically, the problem is quality control: the only conclusive way to test that gametes are functional and can create a viable embryo is to go ahead and do so.

Once gametes can be made, the immediate question will be whether and how embryos derived from them can be created and studied. So iPS research will come full circle, and we will be faced again with the issue that has dogged us since the birth of Louise Brown, the first IVF baby in 1978. Is it moral to use human embryos for research or to create human life with the implicit intention of destroying it during an experiment?

Not conducting this research could cause harm: without it, a pregnancy using an embryo produced by PSCDGs would pose an unacceptable risk for both child and mother. If PSCDGs are to one day join the tools of assisted reproduction, functional tests must be developed to eliminate faulty gametes. Quality tests developed for IVF will be useful in this context and the methods should first be exhaustively assessed on animal models, including non-human primates.

Clearly, appropriate oversight of PSCDG research is needed, but much of the framework for doing so is already in place, in the form of committees for overseeing stem-cell research and institutional review boards for human experimental subjects. Nonetheless, 14 countries already prohibit or restrict PSCDG technology for genetic screening or germline genetic modification, some under penalty of imprisonment. Japan prohibits the creation of human PSCDGs altogether. Where stricter regulations are needed in many countries is on experimental fertility treatments involving human embryos carried out in clinics. But for any country considering legislation in this area, beware of impeding basic research and losing the benefits that could result. ■

# The big ome

It's time to make the case for proteins.

'The body's building blocks', 'cellular machinery': proteins are sometimes stuck with rather mundane labels. Certainly they sit in the scientific shadow of the genome sequence and the eulogies that inspires. So protein biologists face an uphill battle if they are to fire up the research community — and the world beyond — enough to buy into a Human Proteome Project.

The project is just getting past the back-of-the-envelope stage but, in essence, it would systematically catalogue all the proteins manufactured in the body: what they are, where they are and in what abundance (see page 920). A cancer biologist might reveal whether a rogue protein is overexpressed in the tumours she studies compared with levels from healthy tissue that are logged in the proteome register. A geneticist who traces Alzheimer's susceptibility to a region of code could consult the proteome to reveal which proteins are being manufactured from that region in the brain. We can expect this catalogue of proteins to eventually include the targets for almost all future drugs.

There are many obvious parallels with efforts to elucidate the

human genome. A human proteome would be a very expensive and ambitious undertaking and, by its nature, the full benefits cannot be spelt out beforehand. That's how it should be. The fun of the human genome is that spelling out the letters did little to decipher the code — there is so much more complexity than had ever been imagined. We can expect that the proteome will also reveal unexpected delights about the ways in which proteins rally together to perform a task or become tissue.

And yet there was a certain intellectual allure about 'cracking' the human genome that, on the face of it, is lacking from cataloguing all the proteins. And although the human genome is finite, the proteome is almost boundless, because each of the body's proteins may be present in different forms and different amounts in each tissue — and even in precisely the same cell from one moment to the next. That very complexity, however, should inspire, not dissuade.

Proteomic analyses have also been viewed with some scepticism, in part because many studies involving mass-spectrometry profiling have proved difficult to reproduce. And the field has so far largely failed to deliver the disease-tracking biomarkers on which these early efforts were sold. There are few examples of clear, clinically proven benefits. Starting out with this kind of reputation will make a Human Proteome Project particularly hard to sell.

The Human Proteome Organisation (HUPO) has taken some praiseworthy steps towards resolving these issues with, for example, a project to show that different labs can now produce identical results from the same sample. With the rapid evolution of proteomic techniques, the field's reputation and utility is likely to pick up. But HUPO — and the proteomics community — still has a lot to prove and a successful Human Proteome Project is its chance to prove it. It needs to consult widely to devise a strategy that has strong community backing. It will also need to demonstrate that the knowledge stemming from this project will transform the research landscape in the same way that the genome sequence has done.

Failing these, either the project will die or HUPO could risk being left on the sidelines as organizations with money to spend make the decisions about how proteomics should be done.

Times have changed since the Human Genome Project — proposals for mega-biology projects are rather more common and money scarcer. HUPO, and all biologists who love proteins, should articulate clearly and loudly the benefits of sinking US$1 billion into protein biology. It will take much work to get due recognition for the 'cell's work-horses'. ■

# Superconductors *redux*

## Yet another surprise has been uncovered in the complex oxides.

With the discovery of a new class of high-temperature superconductors by reseachers in Japan (see page 922), history seems to be repeating itself. In 1986, Georg Bednorz and Alex Müller of IBM's Zurich research laboratories discovered that a complex oxide of barium, lanthanum and copper became superconducting at 35 K. This sparked an orgy of research that led to the discovery of a related compound (yttrium barium copper oxide) with a superconducting transition temperature of 90 K: high enough to be attained with relatively cheap liquid-nitrogen cooling. It also won Bednorz and Müller a Nobel prize just a year later.

The excitement stemmed from the prospect of exploiting low-cost superconductivity for loss-free electrical transmission, magnetic levitation and other dazzling applications. In superconductors, currents flow essentially without electrical resistance, the source of energy loss through heating. Before 1986, most superconductors were metals and alloys, with generally paltry transition temperatures that no one had managed to push above 23 K.

Now Hideo Hosono of the Tokyo Institute of Technology and his colleagues have shown that another complex oxide, containing lanthanum, iron, arsenic and a little fluorine, will superconduct at 43 K when squeezed by around 40,000 atmospheres pressure: a higher temperature than anything bar the copper oxides (see H. Takahashi *et al*. advance online publication: doi:10.1038/nature06972). Like them, the new material has a sandwich structure of alternate conducting and insulating layers. And like them, doping (replacing some oxygen with fluorine) injects electrons into the conducting layer that contribute to the supercurrent.

The copper-oxide materials have found some uses, but nothing to match the expectations heaped on them in the late 1980s — levitating trains and so forth. It has proved hard to fashion these brittle materials into wires and progress is slow. So cynics might grumble that the new breakthrough will merely renew the same unfulfilled promises.

But the new compound already offers more. For one thing, it reveals how much remains to be discovered about complex solid-state compounds. The combinatorial possibilities for four or more elements are so vast that we have barely scratched the surface, despite efforts to automate the search. And as before, the discovery followed from sound chemical intuition. Bednorz and Müller were led to the copper oxides from the apparently unpromising strontium titanium oxide, a superconductor at a mere 0.3 K, by reasoning what kind of crystal chemistry might boost the requisite interactions between electrons. Hosono and his colleagues similarly picked a systematic path from their initial discovery, in 2006, of superconductivity at about 4 K in a related material, a temperature so low that it attracted little interest. They raised this to 26 K by the start of 2008, and rightly figured that squeezing would take it further.

Most crucially, the 'iron oxypnictides' show that high-temperature conductivity is not the sole preserve of copper oxides. As in that case, superconductivity in the new materials seems to be related to magnetic behaviour. But quite how this works has remained a mystery. With an entirely new family of compounds to play with, the mechanism might be persuaded to start giving up some secrets. With a theory to hand, 'designer superconductors' with much higher transition temperatures might not look like a fool's quest. There could be another Nobel prize in that. For now, it is enough that the oxypnictides have set the community buzzing in a way that recalls the last heyday of superconductors two decades ago. ■

# RESEARCH HIGHLIGHTS

R. PATEL

## Mercuric swallows

*Science* **320,** 335 (2008)

High levels of mercury in rivers present a broader threat to ecosystems than previously thought, according to Daniel Cristol and his colleagues at the College of William and Mary in Williamsburg, Virginia. They found that mercury that entered the headwaters of the Shenandoah River in Virginia between 1930 and 1950 has accumulated not only in aquatic animals and the creatures that eat them, but also in birds that exclusively gobble terrestrial insects and spiders.

Blood samples from 12 out of 13 species of terrestrial-feeding birds breeding near the river contained notably higher levels of mercury than samples the team collected from the same species at uncontaminated sites. But whether aquatic insects venturing onto land or historical floods that deposited the pollutant are responsible for its spread remains unclear.

CONDENSED-MATTER PHYSICS

## Unblinking dots

*J. Am. Chem. Soc.* **130,** 5026–5027 (2008)

Quantum dots are semiconductor nanocrystals that emit light, making them suitable for such things as medical imaging, but their usefulness is undermined by a tendency to blink or even go out when their surface chemistry is disrupted. Encasing quantum dots in a thick alloy shell renders their optical behaviour much more robust, write Jennifer Hollingsworth and her colleagues at Los Alamos National Laboratory in New Mexico.

The team slowly grew many thin layers of cadmium sulphide, zinc sulphide and cadmium zinc sulphide on cadmium selenide quantum dots that were 3–4 nanometres in size until the particles reached 15–20 nanometres across. This altered the colour of the emitted light and increased the fraction of time for which the dots glowed, an improvement that should facilitate new applications in biology and optoelectronics, the researchers say.

R. JACKMAN

PHYSIOLOGY

## Feeling the air

*Cell* **133,** 223–234 (2008)

The process by which mice adapt to low oxygen levels is mediated at least in part by their skin, report Randall Johnson at the University of California, San Diego, and his colleagues.

When they knocked out the ability of mouse skin cells — but no other cells — to make a protein associated with adaptation to hypoxia, the researchers found that an adaptation mediated by the kidneys was greatly reduced. It seems that, normally, when oxygen levels are low, the protein alters blood flow from the skin to the kidneys and encourages the production of erythropoietin (EPO). This tells the body to build red blood cells — a useful response to low oxygen availability.

The work demonstrates a new drug-based mechanism for stimulating EPO production, as well as an intriguing role for skin as an independent environmental oxygen sensor, the authors say.

NANOTECHNOLOGY

## Stencilissimo

*Nano Lett.* doi:10.1021/nl080033t (2008)

Engineers at the University of Cambridge, UK, and the Nokia Research Centre in the same city have stenciled two- and three-dimensional patterns onto the surface of a silicon wafer using silicon oxide nanowires as a mask.

Andrea Ferrari and his co-workers first attached the nanowires to the surface of a silicon wafer with nickel leads. They then etched the wafer chemically, which left extremely thin lines of silicon under the nanowires.

Were the nanowires to be arranged in preset patterns — a task that might be achieved with microfluidic techniques — the etching process would carve nanoscale-circuits into crystalline silicon.

EVOLUTION

## Darwinian chips

*PLoS Biol.* **6,** e85 (2008)

Researchers in California have demonstrated computer-controlled natural selection of an RNA molecule in real time.

Brian Paegel and Gerald Joyce of the Scripps Research Institute in La Jolla took an RNA molecule that can chemically stitch itself to other RNA fragments and reproduce itself as a result. They then loaded this catalyst and some suitable RNA fragments into a chip that can hold minute volumes of liquid, and connected this to a computer. The computer diluted the growing population and resupplied the RNA fragments at ever lower concentrations.

Over time, the system created a selection pressure for more efficient RNA catalysts. After 500 dilutions, the RNA catalyst had accumulated 11 mutations, and performed 90 times faster than the original catalyst.

MICROBIOLOGY

## A misunderstood microbe

*Science* **320,** 382–385 (2008)

The flagellar filaments of *Campylobacter jejuni*, a major cause of bacterial diarrhoea worldwide, are composed of seven protofilaments. The same anatomical

parts contain 11 protofilaments in all other bacteria studied so far.

Edward Egelman of the University of Virginia in Charlottesville and his colleagues uncovered this break from tradition when they examined the species' flagella filaments using electron cryo-microscopy. They propose that *C. jejuni*'s unusual filament structure arose from a minor change in a region of the protein that is responsible for the helical symmetry of flagella filaments.

MOLECULAR BIOLOGY

## Arsenic's action

*Nature Cell Biol.* doi:10.1038/ncb1717 (2008); *Nature Cell Biol.* doi: 10.1038/ncb1716 (2008)

Arsenic trioxide, a poison familiar to readers of Agatha Christie, has been used to treat acute promyelocytic leukaemia since 1992, although how it works has never been fully explained. Two studies have now pin-pointed RNF4, a member of a family of enzymes called ubiquitin ligases, as the reason why the drug destroys PML–RARα, a protein produced in patients with this cancer.

It was already known that arsenic trioxide attaches several copies of SUMO, a signalling molecule, to PML–RARα, and that this addition prompts the destruction of PML–RARα. Ronald Hay of the University of Dundee, UK, Hugues de Thé of the University of Paris and their respective colleagues have shown that the SUMO chains are recognized by RNF4, which directs PML–RARα to an organelle called the proteasome, where PML–RARα is chewed up.

PALAEONTOLOGY

## Amphibious origins

*Proc. Natl Acad. Sci. USA* **105,** 5786–5791 (2008)

Isotopes in the teeth of elephants' extinct relatives are adding weight to the arguments of researchers who think that the Eocene proboscideans from which elephants evolved lived in freshwater.

Alexander Liu at the University of Oxford, UK, and his colleagues measured the oxygen isotopes in enamel taken from the teeth of two late-Eocene genera, *Barytherium* and *Moeritherium*. These genera are closely related 'sister taxa' of modern elephants. The isotopic composition of their teeth matched those of many other aquatic and semi-aquatic mammals, making a case for elephants and sirenians — that is, manatees and dugongs — sharing an amphibious ancestor that lived about 50 million years ago.

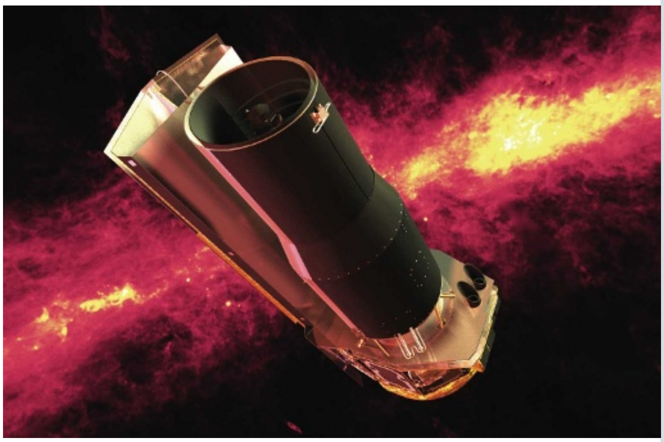
CRYSTALLOGRAPHY

## Probing a crystal

*Phys. Rev. Lett.* **100,** 145502 (2008).

A way to distinguish molecular left- and right-handedness (chirality) has been demonstrated using the SPring-8 synchrotron X-ray source in Harima, Japan.

Yoshikazu Tanaka of the RIKEN SPring-8 Center in Sayo, Japan, and his co-workers have found that the instrument's bright beams of circularly polarized X-rays create disparate diffraction patterns for the two chiral forms of quartz. Normal X-ray diffraction gives identical patterns for both versions. When the angle at which the circularly polarized beam hits the crystals varies between the chiral forms, the intensity at which it is reflected differs. The method can directly ascertain chirality when measuring the rotation of polarized light cannot.



NASA/JPL-CALTECH

ASTRONOMY

## Galactic adolescence

*Astrophys. J.* **677,** 943–956 (2008)

The Spitzer Space Teloscope (picture above) has detected about 2,600 ancient galaxies undergoing a growth spurt. These galaxies are about ten billion light years from Earth and hidden from ground-based telescopes by their own silicate dust, which reddened their light in the same way that our atmosphere reddens sunsets.

Arjun Dey of the National Optical Astronomy Observatory in Tucson, Arizona, and his collaborators used the Spitzer to detect the infrared glow of the galaxies, finding a subset of galaxies that formed hundreds or thousands of stars per year and that had fast-expanding black holes at their centres.

The team believes that this rapid growth lasted for 100–200 million years — a mere blip in galaxy evolution — and that these far-off galaxies went on to resemble the massive galaxies close to our own.

# JOURNAL CLUB

**Vijay Kuchroo**
**Harvard Medical School, Boston, Massachusetts**

**An immunologist muses about inflammation through cell interactions.**

I spend my lab hours trying to understand what prompts T cells — a type of white blood cell — to specialize. Some T cells produce soluble molecules that rattle the immune system into an inflamed state; other cells generate molecules that calm the system back down.

Upon infection, cells such as macrophages — another type of white blood cell — produce soluble molecules called interleukins that direct the fate of the responding T cells. An emerging curiosity in the field is which interleukins make certain T cells become pro-inflammatory, and which cause other T cells to become anti-inflammatory. This decision is crucial for determining whether an immune response induces or suppresses inflammation.

Recently, investigators have turned their attention towards an interleukin known as IL-27. This is produced by activated macrophages and was initially thought to induce IFNγ, a signalling molecule that activates macrophages even more.

But work by Nico Giraldi and his colleagues at Genentech in South San Francisco, and other groups, has recast IL-27 as a molecule that primarily directs T cells to suppress inflammation. In a paper published in March, Giraldi's team confirmed that IL-27 acts in this way because it causes CD4+ and CD8+ T cells to make the anti-inflammatory IL-10, and does not work through an alternative pathway (M. Batten *et al. J. Immunol.* **180,** 2752–2756; 2008). Mice with *Listeria* infections or autoimmune tissue inflammation in their brains and spinal cords generated fewer IL-10-producing T cells when they lacked an IL-27 receptor. Whether an analogous interaction occurs in humans is not known, but, if it does occur, this research could become medically useful.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

# NEWS

## Data show extent of sexism in physics

Women are poorly represented in physics, making up just 10% of faculty in the United States, for example, but the reasons for this have proved contentious. Now a particle physicist claims to have hard data showing institutional sexism at an experiment at one of America's highest-profile physics labs.

Sherry Towers claims that female postdocs worked significantly harder than their male peers but were awarded one-third as many conference presentations proportionally. "There was this shocking difference," says Towers, who now studies statistics at Purdue University in West Lafayette, Indiana. "Particle physics really hasn't moved forward in 30 years."

Towers used data from publicly available work records to chart the careers of 57 postdoctoral researchers, including nine women, who worked on the 'DZero' particle detector at Fermilab in Batavia, Illinois, between 1998 and 2006. Towers herself worked as a postdoc on the project between 2000 and 2005. The findings of her survey were striking, she says. She claims that women did 40% more maintenance work than their male counterparts, and that female postdocs produced significantly more 'internal papers' per year. But based on that productivity they were only one-third as likely to be allocated conference talks as their male peers, she claims (http://arxiv.org/abs/0804.2026).

Conference presentations are critical to a young particle physicist's career. Papers from collaborations such as DZero have hundreds of authors in alphabetical order. Being given the chance to present results at a meeting is a major way for young researchers to stand out. "It's important," says Pauline Gagnon, a physicist with the ATLAS detector at CERN near Geneva, Switzerland. "Being able to give talks is a way of rewarding individuals for their work."

M. JONES

Sherry Towers assessed work data for an experiment at Fermilab.

Most particle detectors have internal committees that allocate conference presentations to researchers. These committees are frequently male-dominated, and Towers believes this lies behind the discrimination. "I don't think for a second that there is a conscious bias going on," she says. But the committees "are in danger of being prone to patronage and cronyism". Male committee members are more likely to nominate male protégés to receive presentation time, she claims.

Some are sceptical of the findings. "I wasn't convinced that the effect she has found is real," says Kevin Pitts, a particle physicist at the University of Illinois at Urbana-Champaign. Internal papers are not necessarily a direct measure of productivity, he argues, and the small number of physicists surveyed is not enough to prove systematic bias. But Pitts is quick to add that he has little doubt that females do suffer gender discrimination: "In fact," he says, "I have personally observed this on more than one occasion."

Female physicists contacted by *Nature* said Towers's data matched their personal experiences of institutional sexism in physics. "You often see a young guy with an older guy gossiping and having coffee, but never a woman," says Freya Blekman, a physicist on the CMS experiment at CERN. "I'm convinced," agrees Gagnon. "There is absolutely no shadow of a doubt in my mind." She says the ATLAS collaboration is thinking about how to address the problem in its own speakers' committee.

After Towers complained, Fermilab launched an internal review in autumn 2006, says Bruce Chrisman, the lab's chief operating officer. An edited copy of the review obtained by *Nature* found that the collaboration "followed its policies correctly". But the investigator, a senior female physicist, added that complaints of gender discrimination in the group "should not be summarily dismissed". There was a general feeling that females were being "passed over" for leadership roles, the report says.

DZero's leaders counter that bias, if it ever existed, is not plaguing the current collaboration. A survey of data between August 2006 and 2007 showed that women gave 17% of all talks despite making up just 12% of the collaboration, says DZero spokesman Dmitri Denisov.

Powers says the investigation didn't focus on postdocs and hasn't led to real changes at DZero. She wants the conference allocation system to be made more transparent and balanced. "The changes that need to be made are simple," she says. "It wouldn't cost them a dime."

And Towers says gender discrimination ultimately forced her out of particle physics. She adds that in 2004 her former employer, a prominent northeastern public university, tried to terminate her contract after she complained that she wasn't given adequate maternity leave. She has since filed a lawsuit against the university. ■

**Geoff Brumfiel**

## Italian group claims to see dark matter — again

**Physicists in Italy claimed last week to have seen particles of dark matter. Their announcement has got their rivals riled and raises questions about what constitutes evidence of a new particle.**

**Rita Bernabei of the National Institute of Nuclear Physics in Rome presented her team's latest results on 16 April at an international meeting of particle physicists in Venice, Italy. Their detector, DAMA/LIBRA (Dark Matter Large Sodium Iodide Bulk for Rare Processes), located deep under the country's Gran Sasso mountain, seems to be observing dark matter, Bernabei says.**

**Most agree that the experiment is picking up something: "They're seeing a signal, there's no doubt about that," says Tim Sumner of Imperial College London. But despite this, critics say that they don't believe the detector has found the elusive particles. "For me, it's not proof that they have seen dark matter," says Gilles Gerbier, a physicist at the Centre for Atomic Energy in Saclay, France. He adds that he's stumped by what's causing the signal.**

**Dark matter is believed by most physicists and astronomers to make up some 85% of the matter in the Universe. Most theories predict that it is some form of massive particle that interacts very rarely — if at all — with regular matter such as atoms. To date, most believe that dark matter has been spotted only indirectly via its pull on rotating galaxies and its effect on the shape of the early cosmos.**

**It's not the first time that Bernabei's team has made this claim. In 2000, they also claimed to have directly observed dark matter. The team uses ultrapure sodium iodide crystals, which theory predicts will give off flashes of light when they are struck by dark-matter particles. After several years of collecting data in the late 1990s, the group saw an increase in the number of flashes**

ANATOMICAL TRAVELOGUE/SPL

**GUT REACTIONS**
Analytical technique shows how metabolism varies between populations.
www.nature.com/news

# Swiss 'dignity' law is threat to plant biology

When it comes to the ethics of experimenting on living subjects, plant biologists have had cause for a certain smugness. But perhaps no longer in Switzerland.

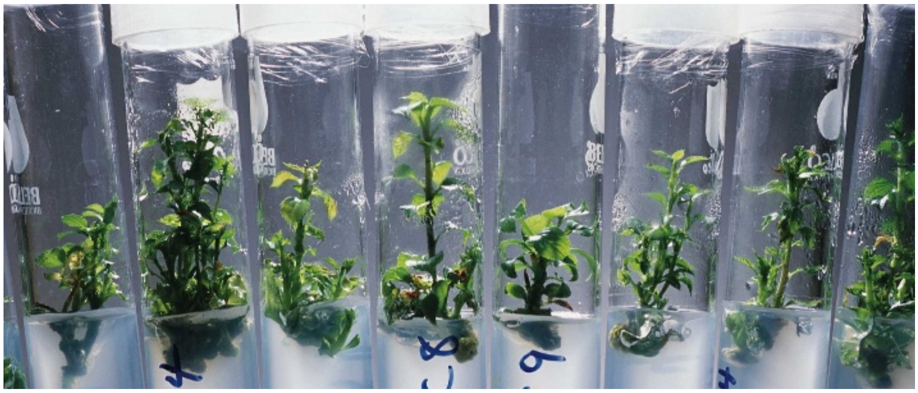
The Swiss federal government's ethics committee on non-human biotechnology has mapped out guidelines to help granting agencies decide which research applications deeply offend the dignity of plants — and hence become unfundable.

Although most people might be bewildered that a discussion on how to define 'plant dignity' should be taking place at all, the stakes for Swiss plant scientists are high. The Gene Technology Law, which came into effect in 2004, stipulates that 'the dignity of creatures' should be considered in any research. The phrase has been widely criticized for its general woolliness, but it indisputably includes plants.

All plant biotechnology grant applications must now include a paragraph explaining the extent to which plant dignity is considered. "But scientists don't know what it means," says Beat Keller of the Institute of Plant Biology at the University of Zurich who is running the first field trial — of disease-resistant corn (maize) — to be approved under the new legislation.

"At the moment not even authorities who decide on grants know what the 'dignity of plants' really means," says Markus Schefer, a constitution lawyer at the University of Basel and a member of the ethics committee. "That's why we were asked to deliberate."

The constitution says that the 'dignity of creatures' must be taken into account in the gene-technology arena, which is why the term has been adopted into the regulations.



J. RICHARDSON/CORBIS

**Living creatures trapped in undignified conditions? A tricky ethical question for Swiss plant biologists.**

The government called on the advice of its ethics committee two years ago to help develop a definition for plants. "My first reaction was — what the heck are we doing considering the dignity of plants," says Schefer. "But this very broad provision exists, and we have to help to prevent a legal mire."

The committee has created a decision tree presenting the different issues that need to be taken into account for each case. But it has come up with few concrete examples of what type of experiment might be considered an unacceptable insult to plant dignity. The committee does not consider that genetic engineering of plants automatically falls into this category, but its majority view holds that it would if the genetic modification caused plants to 'lose their independence' — for example by interfering with their capacity to reproduce. The statement has confused plant geneticists, who point out the contrast with traditional plant-hybridization technologies, for example in roses, which require male sterility, and the commercial development of seedless fruits.

Keller sees the issue as providing another tool for opponents to argue against any form of plant biotechnology, which is already very difficult to conduct in Switzerland. Schefer says that things will start to become clearer when legal challenges to specific research projects come to court, and case law becomes established.

The definition of what constitutes dignity in animals is currently being tested in a Zurich court. Primate-research projects at the ETH Zurich technology institute, which involve separating young marmosets from their mothers, have been put on hold while the court decides if they conflict with the animals' dignity. A ruling is expected this year. Whichever way it falls, the decision is likely to end up in the federal constitutional court. ■

Alison Abbott

---

every June and a decrease during December. The seasonal oscillation, they claimed, was due to Earth's annual motion with and against a stream of dark matter swirling around the Milky Way.

But rival groups failed to see the signal and heaped criticism on the group. They demanded proof that the Italians' detection was not caused by systematic errors such as natural radiation or glitches in their electronics.

This time around, the DAMA/LIBRA team seems to have addressed some of those concerns, according to Bernard Sadoulet, an astrophysicist at the University of California, Berkeley. For example, their new detector uses larger, less strongly radioactive crystals, and they have proven the stability of their detectors.

But once again, DAMA/LIBRA's rivals are coming up empty-handed. Sadoulet's own Cryogenic Dark Matter Search II detector announced a null result earlier this year (see *Nature* 452, 6–7; 2008). And other second- and third-generation experiments are similarly failing to see a signal. If the signal DARMA gets is dark matter, Sadoulet says, "it's certainly not the dark matter we were looking for". He also continues to criticize the group for failing to share enough details of its data. "There has not been enough information provided to the community," he says.

Bernabei defends her group's finding. Most other detectors try to measure direct collisions between a dark-matter particle and an atomic nucleus, she says. DAMA/LIBRA is the only experiment with the years of data needed to see the annual fluctuation in dark-matter particles. And the fluctuation matches "all the several requirements of dark matter", she says, adding that two papers have just been posted on the popular preprint server ArXiv, which provide the details of her work.

But unless another team sees something, Gerbier says, it's unlikely that the wider physics community will accept the Italian claim. "There should be different ways of seeing the same thing," he says. "A single experiment cannot discover dark matter."

Bernabei agrees. Proving her team's findings will take "time and patience", she says. ■

Geoff Brumfiel

# Biologists initiate plan to map human proteome

Ambitious plans to catalogue and characterize all proteins in the human body — a Human Proteome Project — are being drawn up by a small group of researchers. But with a price tag of around US$1 billion, some question whether the organizers can raise enough money or momentum for such an undertaking.

Researchers looked into the idea in the mid-1990s as the Human Genome Project was taking shape — the human proteome seemed a natural successor. However, a coordinated effort to index human proteins never emerged. One reason is that the scale and complexity of the problem proved daunting and nebulous. Protein-coding genes in the body can make tens of different versions of a protein, and each of these can be modified by the addition of chemical groups in countless different ways. All these proteins are being manufactured at differing levels, and at different moments in time, in the 200 or so types of human cell. "It was thought to be beyond comprehension," says John Bergeron of McGill University in Montreal, Canada, former president of the Human Proteome Organisation (HUPO).

Now Bergeron and a group of leading proteomics researchers are putting together a proposal for a large-scale assault on the human proteome. It would reveal which proteins are present in each tissue, where in the cell each of those proteins is located and which other proteins each is interacting with. (The human genetic sequence, by contrast, shows which regions code for proteins but not which are actually making them.) Proponents say that this type of protein catalogue will be invaluable in revealing new drug targets or biomarkers to track the progression of disease. The cost "is absolute peanuts when you consider the importance of mapping the building blocks of life", says Mathias Uhlen at the Royal Institute of Technology in Stockholm, Sweden, who is helping shape the new project.

**"It will be a battle to convince people this is worth funding."**

Two preliminary workshops have been held to discuss the endeavour, most recently in Barbados in January this year. The group plans to consult with the wider proteomics community for the first time at HUPO's Amsterdam World Congress in August.

Those involved in the draft plan say that a human proteome project is now feasible partly because estimates of the number of protein-coding genes have shrunk. It was once thought that there might be around 50,000 or 100,000, but now, just 21,000 or so are thought to exist, making the scale of human proteomics more manageable. And the group plans to focus on only a single protein produced from each gene, rather than its many forms. "We got rid of all this complexity," Bergeron says. "We tried to craft a project that would be doable with easy-to-track milestones."

The plan is to tackle this with three different experimental approaches. One would use mass spectrometry to identify proteins and their quantities in tissue samples; another would generate antibodies to each protein and use these to show its location in tissues and cells; and the third would systematically identify, for each protein, which others it interacts with in protein complexes. The project would also involve a massive bioinformatics effort to ensure that the data could be pooled and accessed, and the production of shared reagents.

Bergeron envisages the work being divvied up between labs around the world. He says that the first stage of the project — which involves amassing existing mass spectrometry proteomic data — would take around six months, and that this would be followed by a pilot project lasting one to three years to do more comprehensive analysis of all the proteins manufactured by chromosome 21, the smallest human chromosome. The whole effort could take a decade.

"It's a huge undertaking for HUPO and they've never done anything like it," says Paul Tempst, an expert in proteomics at the Memorial Sloan-Kettering Cancer Center in New York. The organization has run several large-scale proteomics efforts, such as those to catalogue the proteome in human blood plasma, the liver and the brain. But results from the Human Plasma Proteome project and other proteomics efforts showed that different laboratories — and even the same lab — often identify very different sets of proteins from exactly the same sample. "It told us that there were lots of proteins in plasma and that if you do it in different labs without any effort at technology standardization you get different results," says Tempst.

However, Bergeron and others say that improvements in mass spectrometry techniques have resolved many of the problems of reproducibility that have dogged the field, and that it is now possible to reliably identify a range of proteins if a sample is analysed many times. But an additional difficulty comes in trying to analyse samples containing proteins at wildly differing concentrations, with some present in only a few copies. The three-pronged approach is designed so that shortfalls in one technique can be made up by the others.

Steven Carr, director of proteomics at the Broad Institute in Cambridge, Massachusetts, says there is likely to be broad support for a large-scale proteomics effort, but much debate about how best to do it. Rather than analyse the proteome of one chromosome, he says it may be better to tackle the proteome of mitochondria or the cell membrane because it would reveal more about biology and diseases related to those structures. "It's time to think about something in a systematic fashion — whether this is the project is a different question," he says.

Coordination could also be a challenge.

Heart proteins seen separated in gel.

J. KING-HOLMES/SCIENCE PHOTO LIBRARY

Compared with the human genome project, in which a handful of sequencing centres did the bulk of the work, a human proteome project would involve many more labs, with inevitable issues about data-sharing and competition. "It will be a battle to convince people this is worth funding," says Michael Washburn, director of proteomics at the Stowers Institute for Medical Research in Kansas City, Missouri.

Those involved in the budding project say that they must now muster support from funding organizations such as the US National Institutes of Health and the European Commission — a clear challenge in today's tight funding climate. "It will be a tough sale," says Sudhir Srivastava, head of the Cancer Biomarkers Research Group at the US National Cancer Institute in Rockville, Maryland. Srivastava says that the project may have a better chance of success with a stronger focus on diseases such as cancer. "You need to show clinical utility of the approach before launching a mega-project," he says.

**Helen Pearson**

**See Editorial, page 913.**

**PROTEOMICS**
Find all of *Nature*'s news on this topic in one place.
**http://www.nature.com/news/archive/keyword/proteomics.html**

A. PASIEKA/SPL

# Buckyballs give flash a boost

Flash memory, the workhorse of mobile phones and digital cameras, could be made more efficient by using buckyballs. These spherical fullerene molecules, comprised of 60 carbon atoms, would allow flash memory to operate at a lower voltage and save on power, researchers reported last week.

"We're the first ones trying to borrow molecular electronics concepts and put them into non-volatile memory," says electrical engineer Tuo-Hung Hou of Cornell University in Ithaca, New York, who led the research (T.-H. Hou *et al. Appl. Phys. Lett.* 92, 153109; 2008).

Most desktop computers use a type of random access memory (RAM) that is lost when the power is shut off. Non-volatile memory, however, keeps its content even without power, and flash has become the dominant type. Flash memory holds its zeros and ones in a circuit that contains an island transistor, insulated by a thin layer of silicon oxide.

To write or erase the memory, current is pumped across the barrier, either injecting electrons into the transistor or sucking them out. The charge stays put — most flash memory is guaranteed to last for ten years — but electrons eventually leak through the barrier. The barrier has to be thick enough to prevent leakage, but thin enough for current to pass through during a write or erase.

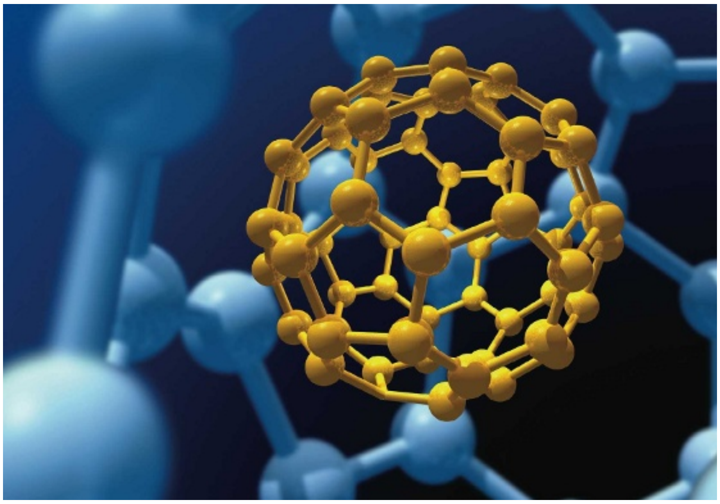
The necessary current is relatively high, which translates to relatively high voltages. Cycle after cycle, the voltage can wear down the flash memory circuits. It also requires peripheral circuitry — which takes up precious space — to boost the low voltages available from batteries or USB ports. And most importantly, it wastes power that could otherwise extend battery life. "The major bottleneck of the current flash memories is the voltage," says Hou.

**"The major bottleneck of the current flash memories is the voltage."**

Enter the buckyball. By adding buckyballs to the barrier layer, the Cornell engineers create resonances that amplify the current during the high-voltage write or erase phase. So the voltage needed during writing or erasing is lowered, by an order of magnitude or more. "It's a very interesting twist," says Sanjay Banerjee, an electrical engineer at the University of Texas at Austin, who was not involved in the research.

Chenming Hu, an electrical engineer at the University of California, Berkeley, says that the concept is attractive, and could help extend the tenure of flash as the dominant form of non-volatile memory. But he cautions that other types are already nipping at the heels of flash, such as magnetic RAM, which stores information using the polarity of ferromagnetic plates, and phase-change RAM, which relies on switching between amorphous and crystalline phases of tiny glass filaments. "The general feeling is something else will replace flash," he says.

**Eric Hand**



Spherical carbon molecules such as these could be used to make flash memory drives less power-hungry.

V. HABBICK VISIONS/SCIENCE PHOTO LIBRARY

# Arsenic heats up iron superconductors

E. HOSHIKO/AP

Could economically viable magnetic levitation be back on track?

A new class of high-temperature superconductors has been discovered in a breakthrough that once again has the scientific community whispering that economically feasible magnetic levitation and lossless energy transmission may be possible.

First discovered in 1911, superconductors can conduct electricity with virtually no resistance and can maintain a current without a voltage being applied. But they initially offered little practical application — most of the metals used in the early work, such as mercury, lead and niobium, had to be cooled to a few degrees above absolute zero before they reached their 'transition temperature', the point at which electrical resistance disappears.

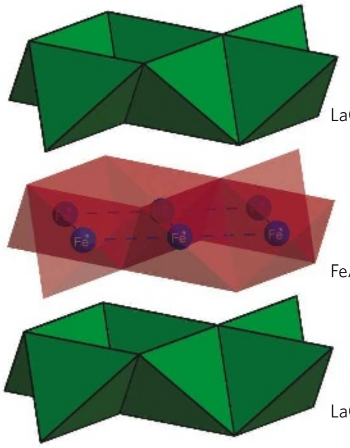
In 1986, scientists identified the first in a family of copper-based, high-temperature superconductors. Within a few years, they had pushed the transition temperature above 100 K — past the boiling point of liquid nitrogen at 79 K, an economically important benchmark. A record of 138 K was set in 1995. But those temperatures eventually stopped rising, dashing hopes that a room-temperature superconductor was possible.

Interest was rekindled two years ago when researchers at the Tokyo Institute of Technology synthesized a new superconductor based on iron rather than copper. The material also featured oxygen, lanthanum and phosphorus, but its transition temperature was just 4 degrees above absolute zero, no better than the very first superconductor discovered a century before.

Then this February, the same group announced an exciting development. The researchers had replaced phosphorus with another pnicogen, arsenic, in the layered material and — boom — the transition temperature shot up to 26 K (Y. Kamihara *et al. J. Am. Chem. Soc.* **130,** 3296–3297; 2008). Subsequent tweaking has already boosted that temperature above 50 K. "We all were surprised," says materials scientist Hideo Hosono, who led the study.

The superconductor community is now buzzing. A number of groups around the world are working with the iron-based material, ramping up efforts and trying other pnicogens such as bismuth. Every week, scientists are posting papers on the preprint server arXiv with new proclamations of the material's properties. "It's phenomenal, because we've broken the tyranny of copper," says Paul Canfield, a physicist at the Ames Laboratory in Iowa.

The iron-based family might provide a fresh opportunity to engineer superconductors that operate at practical temperatures. It also offers chemists a chance to finally figure out how high-temperature superconductors work.




**A hot sandwich: the superconducting iron and arsenic filling in this crystal is fed electrons from the layers above and below.**

Conventional low-temperature superconductors were explained in a Nobel prizewinning theory developed by John Bardeen, Leon Cooper and Robert Schreiffer. Vibrations in the metallic crystal lattice squeeze electrons into pairs, overcoming their mutual repulsion. The paired electrons, operating in a single quantum state, move freely without resistance. A secondary property of superconductors is the blocking out of magnetic fields, which leads to their ability to levitate magnets.

The discovery of the copper-based family of ceramic superconductors raised interesting questions that are still unanswered. "We cannot make sense of it with the theoretical physics we know," says Jan Zaanen, a condensed-matter theorist at the University of Leiden in the Netherlands. "We're dealing with a big mystery." Theorists strongly suspect that magnetism is involved in high-temperature superconductivity. In certain types of metal, the movement of electrons can become 'jammed', and the metal becomes an insulator. As the electrons sit in place, they spin, forming tiny magnets — but with opposite polarities that cancel out the net field. Such materials are said to be antiferromagnetic.

Like the copper-based superconductors, the new iron-based ones seem to be antiferromagnetic. They also share a planar structure, where the superconducting 'action' occurs in the middle of the chemical sandwich (see diagram). As magnetism can exist at very high temperatures, the hope, barely expressed because it's so audacious, is that an understanding of the mechanism could lead to a way of designing custom superconductors, ones that work at room temperature. "Maybe, maybe, maybe," says Hosono.

Much work remains. The researchers have so far worked only with superconducting powders and need to grow crystals, which are more useful for experiments. Copper-based superconductors are easy to work with, but the new family involves a trickier chemistry, and there is the toxicity and volatility of arsenic to consider. Last month, a sealed tube of arsenic exploded at the Argonne National Laboratory in Illinois, when a research group started to synthesize the new material for the first time. No one was hurt, but some academics won't work with arsenic because they don't want their graduate students to take risks.

However, any worries or doubts are, right now, outweighed by excitement and renewed enthusiasm for high-temperature superconductivity. "It's early days," says Canfield, "but we've already doubled the transition temperature." ■

**Eric Hand**

**See Editorial, page 914.**

# Politically correct names given to flu viruses

Names of flu viruses appearing in the scientific literature have undergone something of a mutation this year. The group of H5N1 avian flu viruses, once known as 'Fujian-like', has morphed into 'Clade 2.3.4'. And its 'Qinghai-like' cousin is now called 'Clade 2.2'. They are part of a complete revision of H5N1 nomenclature recommended by the World Health Organization (WHO).

The system is now more politically correct. It avoids the "stigmatizing labelling of clades by geographical reference", according to the WHO. In 2006, when scientists assigned the name 'Fujian-like virus' to a vaccine-resistant strain of H5N1 that spread across Asia, China objected strongly to the name. Government officials argued that the southeastern Chinese province of Fujian was tainted by association as the virus spread.

But a WHO official told *Nature* that discussions for a revised nomenclature were underway before the spat, and that the "impetus was scientific". "We recognized that naming based on geography was probably not a good idea for other reasons," she says.

Flu experts say that the new naming recommendations, which have been posted on the WHO website and have been submitted for publication in the journal *Emerging Infectious Diseases*, make a lot of sense. They replace an ad-hoc system, in which researchers would invent their own names for newly isolated groups, with a systematic one in which names have greater biological relevance.

H5N1 viruses are classified on the basis of their haemagglutinin genetic sequence into ten phylogenetic clades — distinct branches of the virus's family tree — and the new system names the viral subgroups according to their position in the tree. It aims to be more logical. For example, Clade 2.3.4 viruses are not restricted to Fujian province, they have caused cases in humans in China, Laos, Myanmar and Vietnam. Clade 2.1 viruses dominate in poultry and humans in Indonesia, for example, whereas Clade 2.2 viruses have the largest geographical spread, causing outbreaks in more than 60 countries from Azerbaijan to Nigeria and Pakistan.

"I find the geographical naming system rather confusing and unspecific; this more precise numbering system is far more rigorous," says Edward Holmes, a flu genomicist at Pennsylvania State University in University Park. The new system does not affect the naming of individual H5N1 virus sequences — held in the GenBank database — that make up such clades. These will continue to display geographic and species data on the sample. Geographic information is relevant, says the WHO official, because it locates the sample's place of origin.

But experts say that the location data in GenBank flu names is woefully inadequate for serious spatial science, such as analysing the spread of virus genotypes or how spread relates to other spatial data on land use, climate and trade. "There's nothing more frustrating and useless to a phylogeographer than a perfectly sequenced isolate labelled 'China'," says Rob Wallace, an evolutionary ecologist at the University of California, Irvine. "With Global Positioning System (GPS) coordinates and fine enough sampling we would really be able to tease out how the virus evolves and spreads across farms and wetlands."

M. EARTHY/AP

Bird flu knows no national borders.

Indeed, not one H5N1 sequence in GenBank currently contains GPS coordinates, often because those submitting sequences have no use for such data. "Given the importance of avian flu, I've been amazed to find out how difficult it has been to get information on the geographical distribution of essential variables," says ecologist Marius Gilbert, of the Free University of Brussels. The WHO official agrees that the current sequence-naming system is not as useful as GPS coordinates, but says that "the WHO doesn't see its business as making that sort of requirement".

GenBank only introduced fields for latitude and longitude in submitted entries in 2005 and few scientists have so far used them, except in spatially aware fields, such as forest entomology and biodiversity. ■

**Declan Butler**

ACTION PRESS/REX FEATURES

## SHOWBIZ NEWS

**Formulaic pop**

R&B diva Mariah Carey has somewhat surprisingly titled her latest album E=MC², not as a homage to the main man Einstein, but because it's allegedly short for "Emancipation equals Mariah Carey times two". Somebody please teach her some maths before she releases more music.

## ON THE RECORD

**"Contrary to recent press reports, NASA offices involved in near-Earth object research were not contacted and have had no correspondence with a young German student."**

The agency denies accepting help from 13-year-old Nico Marquardt in working out how likely we are to be hit by the asteroid Apophis.

## WORDWATCH

**Nerdic**

Also known as 'geek speak', it is the fastest-growing 'language' in the world, adding about 100 new terms every year, including the verb 'to Google'.

## SCORECARD

▲ **Moon crops**

European Space Agency scientists find that plants can grow, without needing extra food, in crushed-up rock very like the lunar surface.

▼ **Space tourism**

Just as everyone else jumps on the space-tourism bandwagon, the world's top space-travel company, the Russian government, announces that it may stop selling space voyages in 2010, to save more seats for cosmonauts.

*Sources: Mariah Daily Journal, NASA, Ananova.com, BBC, The China Post*

SIDELINES

NATURE|Vol 452|24 April 2008

Q&A

# Fake drugs: lessons for the world

**Dora Akunyili** is director-general of Nigeria's National Agency for Food and Drug Administration and Control (NAFDAC). Since her appointment in 2001, she has led a successful crusade against counterfeit pharmaceuticals, which are responsible for millions of deaths worldwide each year. *Nature* caught up with her last week in Washington DC.

**What is top of your agenda at the moment?**
I want to bring the incidence of fake drugs in Nigeria down to single-digit levels. And I want to institutionalize our processes and systems at NAFDAC so that it is as strong as food- and drug-regulatory agencies in developed countries. That would not only save lives but would help our food and drug industries to capture the African market.

**Since you took office the proportion of fake drugs in Nigeria has fallen from some 41% to around 10%. What did you do?**
We achieved that through restructuring NAFDAC, repositioning staff for better effectiveness, and public enlightenment campaigns — alerting people to the fact that there was a problem. We also started focusing on those countries that supply these drugs to Nigeria, especially India and China. We put in place Indian and Chinese nationals as independent analysts in these countries to recertify drugs before they are imported into Nigeria. The drug importers pay them.

We also go anywhere in the world to inspect facilities producing drugs, food, cosmetics or any NAFDAC-regulated product before we register the product. And we insist on preshipment information before drugs are accepted in Nigeria. We must get all information, from which drug is coming in what quantity to the final detail of the conveyance vessel. And before the financial import document is processed by Nigerian banks, they must get NAFDAC clearance. Above all, we insist on a certificate of free sale. This states that before a drug is registered in Nigeria it must already be used in the country of production.

**What about the drugs that slip through?**
We mop up what is already in circulation by sustained, systematic surveillance. We depend on reports from the public, from patients, from hospitals, even from journalists. In 2003, it was Nigerian journalists who reported the deaths of children during open-heart surgery in a teaching hospital. We investigated and found that the drugs used during the surgery were substandard. In this mopping up exercise of fake products, we have closed three different drug markets and sealed up shops.

P. HOLDEN PHOTOGRAPHY/AMERICAN ENTERPRISE INST.

**Has the problem simply moved elsewhere?**
When Nigeria got too hot for drug counterfeiters, they started migrating to other West African countries and became a big problem to them. We instituted the West African Drug Regulatory Authorities Network as a platform to work with other countries to encourage them in this fight. We are also occupying an important position in the International Medical Products Anti-Counterfeiting Taskforce, for which I am the vice-chair.

**Do you get support from the Indian and Chinese authorities?**
They are forced to support us. I don't think they actually wanted to support us because fighting these counterfeit drugs weakens their economy. That is why they have strong regulations for drugs meant for domestic consumption and little or no regulation for drugs meant for export. We have banned 30 Indian and Chinese companies and one Pakistani company from importing drugs into Nigeria. We have better support from India than from China.

**Do you think that the execution last year of the head of the Chinese drug-regulatory agency was an effective measure?**
It was not effective. It was a face-saving measure. It was because America had spoken. If the Chinese government is serious, then who are the workers that worked with the man? Who facilitated what he did? No chief regulatory officer could do what he did alone. There are people farther down the ladder. We are still waiting for the Chinese government to tell us who they are and what they have done to them.

I believe that another reason they killed him is probably because Chinese children died. If all of those drugs were for export, they wouldn't have killed him. It was because their own people had been affected.

**You were the subject of an assassination attempt five years ago. How safe do you feel nowadays?**
Since 2003 I have, happily, not been shot again. But I still experience threats here and there and I don't think they have given up. But my security has been beefed up so much that it is getting more difficult for them. Right now I have at least eight policemen around me all the time and they are well-armed.

**Why has nobody been convicted yet?**
It is a very, very unfortunate situation. After the shooting, the government was able to get at the people who organized it because before it happened they were boasting that after December 2003 there would be no NAFDAC, and that I would not live to see 2004. So one arrest led to another. The case started in 2004 and lasted over one and a half years with 58 exhibits and 19 witnesses. And after all that, the judge's verdict was that he had no jurisdiction to hear the case. The case was sent to appeal court. The appeal court ruled against the judge and said: 'you must hear it, you have jurisdiction'. That was early last year. And for over 12 months there has been adjournment on that case, after it was sent back to the high court. Even last week we were in court and it was adjourned for next week.

**Are fake drugs an issue that people in Washington should worry about?**
The problem is already here in America. And it is one that can swallow the whole human race if nothing is done. Resistant strains of microorganisms caused by substandard antibiotics do not need visas to travel from country to country. ■

**Interview by Meredith Wadman.**

## Bubble-fusion engineer sues other scientists

Purdue University in West Lafayette, Indiana, has sent federal authorities a report of its investigation into the work of nuclear engineer Rusi Taleyarkhan, *Nature* has learned.

Taleyarkhan is known for his claims to be able to produce fusion by collapsing bubbles in deuterated acetone by bombarding them with sound waves. His work had been subject to allegations of research misconduct, which Purdue had been investigating.

Taleyarkhan's lawyer, John Lewis of Lewis and Wilkins in Indianapolis, says that the final decision now lies with the inspector-general of the Office of Naval Research in Arlington, Virginia. The office funded some of Taleyarkhan's research through grants. The Purdue report is confidential, but Taleyarkhan plans to contact the office to challenge the report's "accuracy, adequacy and completeness", says Lewis.

The engineer is also pursuing a defamation lawsuit, filed last month in Tippecanoe County in Indiana against scientists who questioned his claims of bubble fusion. The defendants include Purdue faculty members named in *Nature* articles, which helped trigger an institutional review. "The complaint filed is to clear up my reputation," Taleyarkhan says.

## US biodefence agency appoints first director

The body set up last year to coordinate the US government's biodefence efforts has picked its first permanent director.

The Biomedical Advanced Research and Development Agency (BARDA) is to be headed by Robin Robinson, the Department of Health and Human Services announced last week. Robinson is currently deputy director of BARDA's influenza and emerging disease programme, and has overseen the development and stockpiling of vaccines against the H5N1 avian flu virus.

B. SKLAR/AP


Anthrax scares in 2001 prompted Congress to boost US biodefence efforts.

## Botanical art gallery opens in London

This image of the flowering house plant *Medinilla magnifica* by artist Margaret Ann Eden is now on display at the Royal Botanic Gardens at Kew in London.

It is part of an exhibition in the world's first year-round gallery dedicated to botanical art. The exhibition features works from Kew Gardens' 200,000-piece collection as well as from Shirley Sherwood, who has collected images from around 200 artists. The gardens' officials say that the Sherwood collection is possibly the most important private collection of modern botanical art in the world.

M. A. EDEN

BARDA oversees Project BioShield, a $5.6-billion effort to purchase biodefence countermeasures that has been criticized as a giveaway to industry (see *Nature* 431, 1023; 2004). It is the culmination of much congressional discussion over what to do about biodefence after the still-unsolved postal anthrax attacks of 2001.

## Research assessment panel advised to destroy all notes

All personal notes made by those conducting a massive assessment of university research in the United Kingdom should be destroyed, the organization in charge of the exercise has recommended.

Last week, the *Times Higher Education* published extracts from a letter written by Ed Hughes, head of the team managing the Research Assessment Exercise (RAE), which determines funding allocations for university departments. The letter recommended that members of assessment panels destroy all personal notes relating to their decisions.

However, the Higher Education Funding Council for England, which distributes the English portion of RAE-related funds, says that minutes of assessment meetings will still be published so that decisions can be scrutinized. The destruction of personal notes, the council says, is to prevent panel members being over-burdened with requests to publicly release their notes.

## GlaxoSmithKline does deal to develop microRNA drugs

Pharmaceutical giant GlaxoSmithKline has negotiated a deal to develop drugs aimed at four microRNA molecules that have been linked to inflammatory diseases such as rheumatoid arthritis and inflammatory bowel disease. The agreement, with Regulus Therapeutics in Carlsbad, California, marks the first time a large pharmaceutical company has announced a collaboration that is exclusively focused on microRNAs, which silence genes by intercepting messenger RNA molecules.

Unlike small interfering RNAs that also work in this way, microRNAs usually downregulate expression of many genes at once. They have thus been associated with diseases involving multiple changes in gene expression, including cancer and viral infections.

Regulus will initially earn US$20 million from the alliance, and then receive up to $144.5 million more for each of the four microRNA targets for which it meets drug-development milestones.

## UN ruling makes Australia an even bigger country

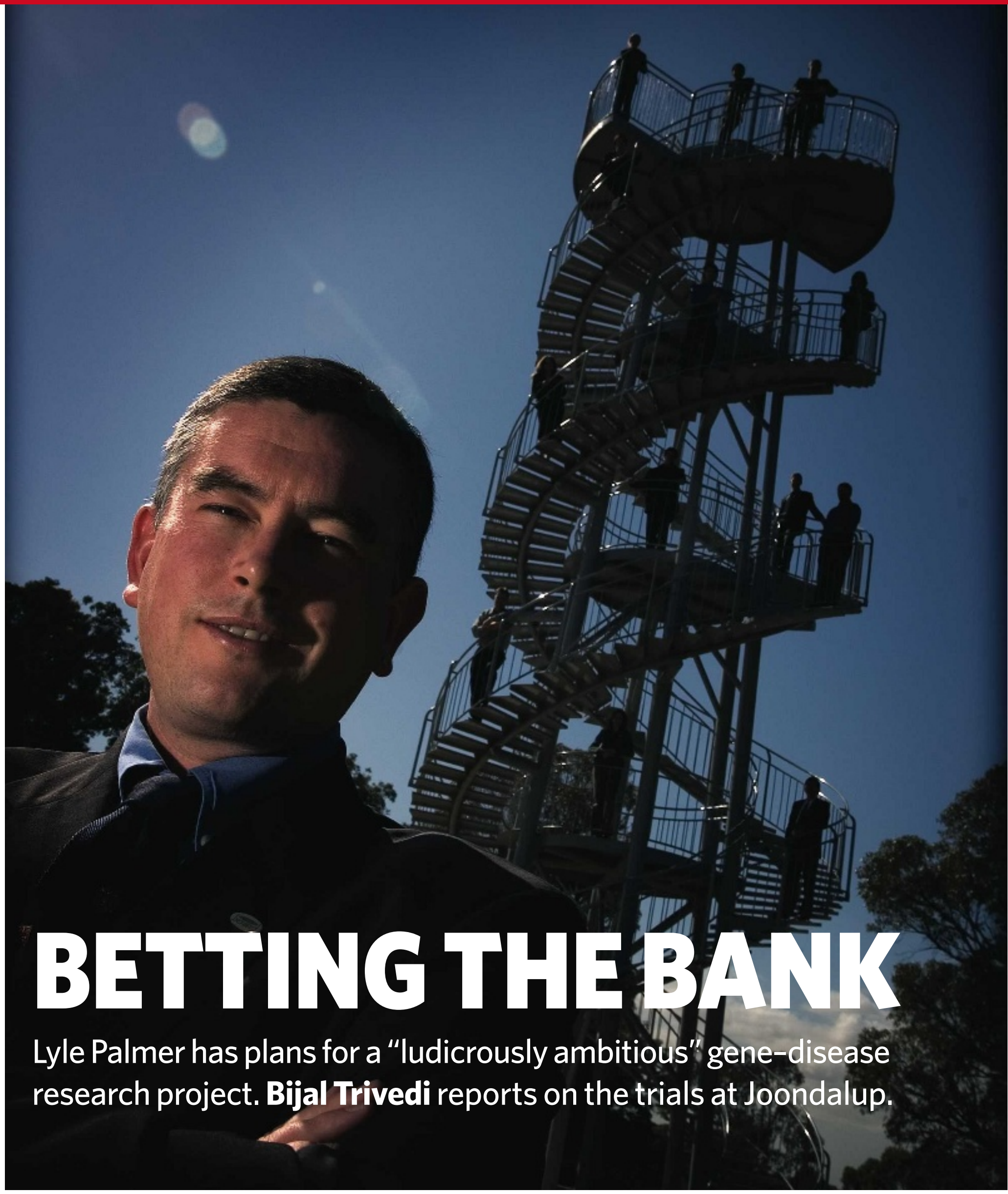
Australia has become the first country to successfully claim an extension of its rights to the sea floor.

Resources and energy minister Martin Ferguson announced on Monday that the United Nations had extended the area of the country's continental shelf by 2.5 million square kilometres. "I'm pleased to announce that Australia, the largest island in the world, has been dramatically increased in size," he said.

The country's rights have been extended beyond the usual limit of 200 nautical miles (370 kilometres) in a number of areas, including those off the main Australian coast and a number of smaller islands. This gives Australia exclusive rights to exploit any resources on the sea floor within this area, including oil and gas.

"It opens up a wealth of exploration entitlements," says Ferguson. "This is potentially a bonanza."

Other claims under the United Nations Convention on the Law of the Sea have been made by countries including Mexico, France and Great Britain. Contentious claims over the sea floor of the Arctic are expected from Russia, Canada and Norway.



# BETTING THE BANK

Lyle Palmer has plans for a "ludicrously ambitious" gene–disease research project. **Bijal Trivedi** reports on the trials at Joondalup.

A. FRANK/RAW IMAGE

Every hour of Connie Colgan's day is carefully choreographed. Rising at 5:30 a.m. she prepares for a marathon of waking, dressing and feeding her six children before chauffeuring the eldest four to school. Sean and Conor, 3 years old and 19-months respectively, wreak havoc at home as their mother tidies, does laundry, cooks and then ventures out again to ferry the kids to swimming, Irish dance, piano and gymnastics classes. Gruelling as her schedule is, Colgan is adding another routine — she has enrolled her family in an ambitious new health study that will rank her brood among the most biologically characterized humans on Earth.

After donating blood samples, from which DNA will be extracted and probed at one million locations with gene chips, Colgan's family, and potentially 80,000 other Western Australians, will enter a newly built facility where they will wind their way through a maze of tests: whole body scans, retinal exams, hearing tests, muscle strength, respiratory tests and more. In all, about 3,500 measurements will be made of each participant. Then, every three years, they'll do it again.

The effort, planned to start next year, is one of the newest in a series of population databases, or biobanks. By storing DNA and blood samples as well as the medical and family history of each volunteer, biobanks provide a tremendous resource from which researchers have been able to extract risk factors for common diseases — such as obesity, asthma, depression and heart disease — that are deluging the healthcare systems of developed and developing nations. Genomewide association studies and more complex genome interaction analyses have begun to reveal ways in which these diseases might be treated.

Beyond probing genes, however, the more ambitious of these biobank projects hope to account for the effects of diet and the environment and even such factors as state of mind. With these interactions becoming cheaper and more realistic to peruse, biobanking efforts have sprung up all over the world. They can be pricey, difficult to develop, ethically complex to navigate and their returns are only speculative. But none of this has deterred Lyle Palmer, the brains behind the efforts in Western Australia, from devising one of the most ambitious and cutting-edge biobank projects to date.

> "How about we look at every single body system and take every single measurement?"
> — Lyle Palmer

For Palmer, chair of genetic epidemiology at the University of Western Australia in Perth, the perfect site for what he calls a "ludicrously ambitious" project is Joondalup — a regional city north of Perth — where efforts to recruit participants and build excitement about the Joondalup Family Health Study have been ongoing since 2005.

Palmer (pictured above in front of the 'DNA Tower' in King's Park, Perth) speaks quickly, especially when excited. And the prospects for epidemiological research in Western Australia excite him. The public-health system is one factor. It tracks the prescription-drug use of all residents. Plus, roughly 40 years ago the Western Australia state government passed a law that all public and private hospitals would share patient records with the health department. The data bank contains births, deaths, marriage registrations, emergency-department diagnoses, surgical history, and midwife, mental-health and cancer records. "I don't know why they made that decision but it is a godsend that they did. And we are now the beneficiaries," says Palmer.

Drug data and medical history of all the state's residents will provide a welcome context for the imminent flood of genetic data. Although the state lacks the centuries of genealogy that help power biobanks in Iceland or Utah, Palmer and his colleague John Bass, who studies health informatics at Curtin University of Technology in Perth, are cobbling together a genealogy of the state's residents that dates back to 1840.

## Line of volunteers

Joondalup's signature will be the scope of its medical testing. In planning the project, Palmer is pushing to not just build on other prospective epidemiology studies, but to blow them out of the water. "How about we just look at every single body system we can think of and take every single measurement?" he says.

Once the project is in full swing, Palmer's colleague Anne Pratt, who is leading the data collection, envisions that a new volunteer will enter the testing facility every 15 minutes. After a quick interview, questionnaire, and collection of blood, urine and saliva, each person will embark on a colour-coded 'throughput cycle' that will guide them through various tests. Among the tests they might take are

dual-energy X-ray absorptiometry scans showing fat and body mass, brain positron-emission tomography scans, multi-slice computed-tomography scans, cardiorespiratory stress tests, hearing tests, and measurements of back muscles endurance, lung volume and bronchial responsiveness, more than 20 lengths, widths and circumferences, more than 50 blood chemicals and thousands of other variables.

The University of Western Australia recruited Palmer from Harvard in 2003 to build and lead a world-class genetic epidemiology facility that would capitalize on the state's untouched genetic resources. He admits that "this wasn't the most attractive academic job or the most attractive from a financial point of view". But the tug of family, he says, and the desire to give back to the system that had given him a free education, led him back to Australia. And, if Palmer were going to leave Harvard, he was determined to create something that would rival all other epidemiology resources.

> "We found more genes for complex diseases in 2007 than in the entire history of the field."
> — Lon Cardon

Palmer describes his plan with the uncompromising confidence of a luxury-car salesman with no need to oversell. He should be practised. He has spent the past four years explaining to politicians, scientists and companies how troves of untapped medical data could be harnessed to make Western Australia the world's go-to region for gene hunting, drug testing and health research. The project has secured Aus$150 million (US$140 million) from local institutions and other collaborators that should carry the project for three years.

## Battle of the banks

In the past 18 months, there has been something of a revolution in gene hunting business. "We found more genes for complex diseases in 2007 than in the entire history of the field," says Lon Cardon, a statistical geneticist and the newly appointed leader of the genetics division at GlaxoSmithKline (GSK).

The discovery boom began as the quantity of genomic data hit a tipping point. After the human genome project was completed in 2001, another large project began to generate a map of human genetic variation by cataloguing single nucleotide polymorphisms (SNPs). This, in turn, laid the groundwork for the rise of genomewide associate studies that were geared to look at the frequencies of these SNPs in disease populations and make it easier to link genes to diseases. Biobanking is the next logical step to translate genetic data into clinical applications, says Cardon.

At least 18 countries have launched or are planning population biobanks including Iceland, the United Kingdom, Estonia, India, Sweden, China, Mexico, Japan, Gambia, Canada and the United States. Pretty much any tissue collection is a biobank, and they vary wildly. Some are cohorts in which the volunteers provide a DNA sample and are followed up over time. Others are snapshots, surveys with a single sample and questionnaire but no follow-up. Sizes of biobanks range from just a few thousand to half a million recruits.

DeCODE Genetics of Reykjavik, Iceland, has become synonymous with biobanking and gene hunting. It recruits participants through Icelandic physicians. Volunteers give blood, from which DNA is extracted, and the doctors share the diagnoses with deCODE. With this approach deCODE doesn't characterize the physical, health or behavioral qualities — the patients' 'phenotype' as opposed to genotype — but rather relies on the physicians' measurements. To date, the company has collected DNA and blood samples from 120,000 Icelandic citizens; 95% agreed to allow deCODE to use their DNA for any study approved by the national bioethics committee.

The UK Biobank in Stockport, one of the largest planned biobanks, intends to recruit 500,000 adults aged 40–69 years. When volunteers visit an assessment centre each one gives written consent, completes a lifestyle questionnaire, enters a one-on-one interview to provide medical history, and then undergoes a medical examination and sample collection.

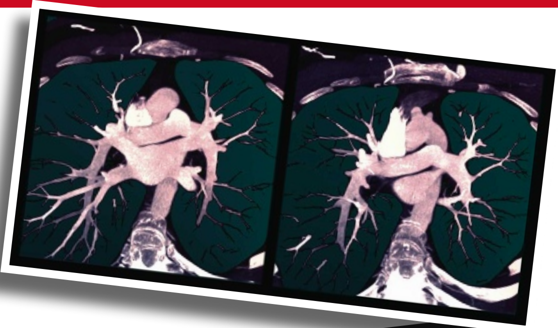
Roughly 20 months ago the Children's Hospital of Philadelphia in Pennsylvania launched the Children's DNA Database. The goal: to collect the DNA of 100,000 of its child patients and scan for genetic markers associated with cancer, irritable bowel disorder, epilepsy or diabetes, for example. After just a year and a half, the hospital and its 29 satellite medical centres have amassed 48,000 blood samples with approximately 35,000 from children and 13,000 from parents.

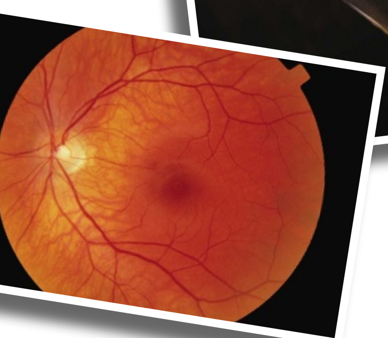
"It has been very successful," says Hakon Hakonarson, the director of the Center for Applied Genomics at the hospital and ex-deCODE executive. The response rate is high — only 10% of parents decline, and the DNA bank only requires a one-time contact with the family to collect blood. But if the parents agree, researchers can access the child's electronic medical records, which are updated every time the child visits the hospital or its satellites.

Officials at the US National Institutes of

FROM TOP: BSIP, GONDELON/SPL; N. MORRISON/SPL; T. BEDDOW/SPL; ZEPHYR/SPL; E. ECKSTEIN/PHOTOLIBRARY; E. HILDEBRANDT/SPL; P. PARKER/SPL; P. GOETGHELUCK/SPL

NATURE|Vol 452|24 April 2008





Health (NIH) in Bethesda, Maryland, have also argued for the establishment of a 500,000 person biobank (see *Nature* **429,** 475–477; 2004). Some disagreed, and a number of factors have stood in the way, including the lack of a national healthcare system and linked data. Estimates have pegged the price tag of such a venture at around US$3 billion dollars, a figure that doesn't look likely to come from NIH's tight budget.

## Regulation and collaboration

The international consortium Public Population Project in Genomics (P3G) keeps tabs on 123 biobanks, mostly cohort studies, and tries to ensure that members follow basic legal and ethical guidelines. One of their major objectives is to encourage collaboration between biobanks and to foster data harmonization by standardizing questionnaires and research protocols. Such collaboration, however, has its downside. "I find many of the cohorts designed today are very similar — same choices of phenotype," says Tom Hudson, scientific director of both P3G and the Ontario Institute for Cancer Research in Toronto. "Working too closely seems to have suppressed innovation."

> **"The people who contribute data to our discoveries have tangible benefits coming from it."**
> — Kari Stefansson

In this respect, isolation may have served Palmer well. Hudson, who was a visiting professor at the University of Western Australia two years ago, says he was struck by the diversity and novelty of phenotypes that Palmer's team will be measuring. "That's what impressed me."

In addition to all the physical measurements, DNA and blood analysis, Palmer is also collecting data about lifestyle — health, behaviour, work, family, school and community — using Internet questionnaires that probe subjects as wide ranging as computer use, bullying, noise exposure and Australian values.

Kari Stefansson, the chief executive of deCODE Genetics, calls Palmer's "deliberate brute-force phenotyping" excessive. But no one has optimized the biobank formula, yet. "It is so early in the knowledge of the genome that we have to cast a wide net because we don't know what we are going to catch," says Teri Manolio, director of population genomics at the National Human Genome Research Institute at the NIH. "One of the challenges with a study like this is not picking what to include but what to leave out."

DeCODE certainly leads the pack when it comes to linking genes, gene variations and loci to specific complex diseases. But its ethical blunders, say many in the field, have more demonstrated how not to launch a biobank. In 1998, the Icelandic government granted deCODE a 12-year exclusive licence to a central database containing the health records of all Icelanders. This, in combination with 1,000 years of genealogy and DNA from the Icelandic populations, would have proved a formidable tool for hunting disease-causing genes. The Icelandic Supreme Court overturned this decision in 2003 because such "presumed consent" rather than informed consent was unconstitutional. In 1998, the drug firm Hoffmann LaRoche, based in Basel, Switzerland, struck a US$200 million deal with deCODE for the right to develop drugs based on deCODE's data. But the perception of a foreign company profiting from Iceland's medical database turned the stomach of many citizens. The company tried to calm the waters by promising free drugs and diagnostics to Icelanders.

"We feel very strongly still today that the people who contribute data to our discoveries have some tangible benefits coming from it," says Stefansson.

Other biobank efforts have steered clear of these informed-consent troubles. Few, if any, promise free drugs or diagnostics, fearing, as many bioethicists have pointed out, that such lures coerce sick individuals and their families. Palmer who says that deCODE began with a "crazy model" insists that the Joondalup Family Health Study has taken adequate measures to protect its participants.

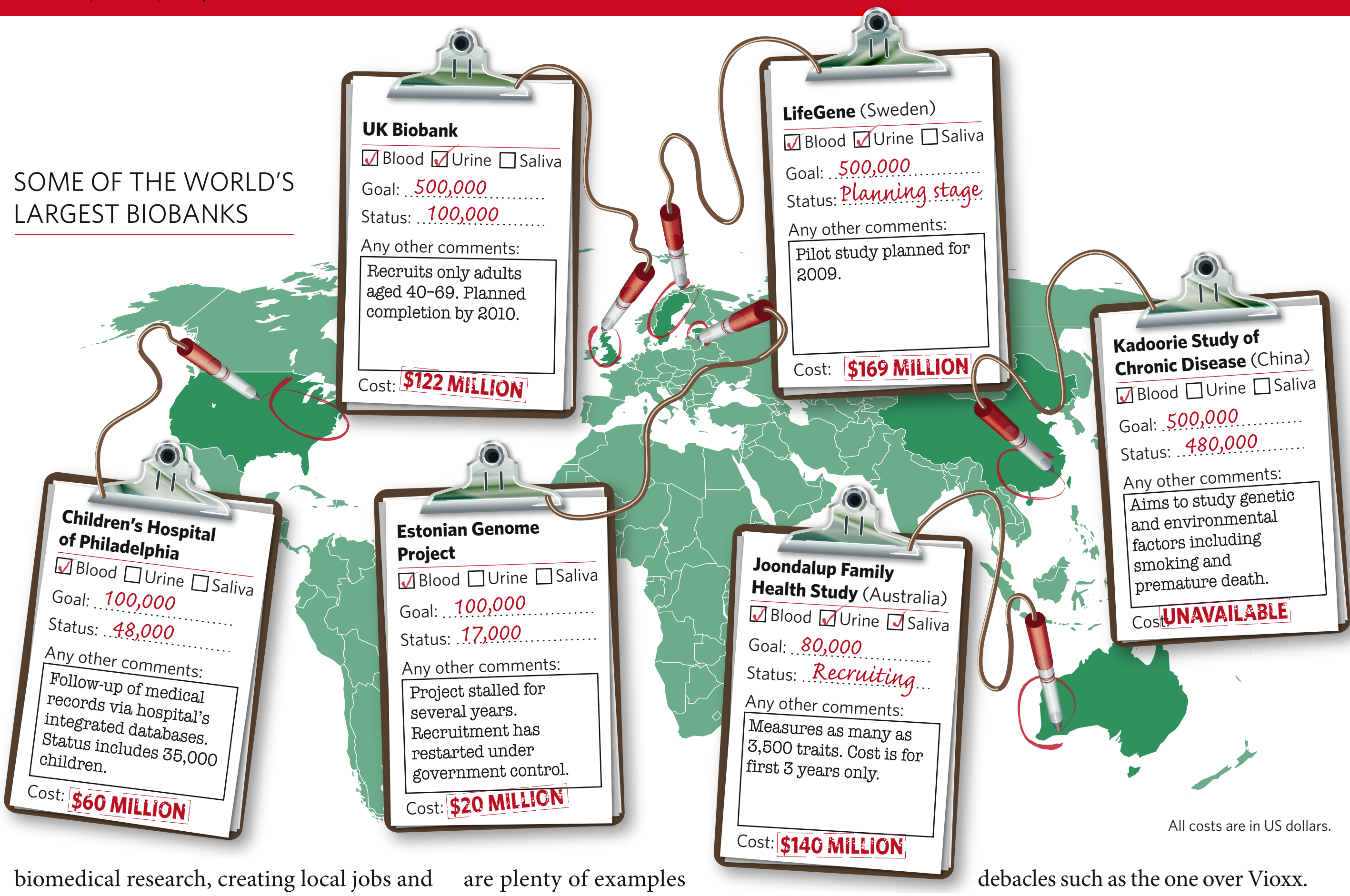
Despite these reassurances, Michelle Kosky, executive director of the Health Consumer's Council of Western Australia, says that the participants are still vulnerable to human-rights violations. She is concerned that the database containing genetic data could be hacked, jeopardizing individual rights to income insurance protection and life insurance. And consent continues to plague many biobanks that include children, including the Joondalup study. "When children turn into legal adults, they don't have a say how their DNA and medical records are used — they're gone, they're already out there," says Patricia Roche, a bioethicist at Boston University School of Public Health in Massachusetts. "Do parents really have the right to do that?"

## If you build it, will they come?

"Change the world" is the promotional motto emblazoned on the orange rubber bracelets that Joondalup-study participants wear, reflecting Palmer's belief that the discoveries will improve health around the globe. And, just as deCODE put Iceland on the map for

ZEPHYR/SPL; S. FRASER/SPL; BSIP, GONDELON/SPL; C. JAY/SPL; T BEDDOW / SPL; P. PSAILA/SPL; P. PARKER/SPL ; M. MASLO/SPL

SOME OF THE WORLD'S LARGEST BIOBANKS




All costs are in US dollars.

biomedical research, creating local jobs and opportunities, he expects that the Joondalup study will do the same for Western Australia. He plans to do this by holding the data in what he calls a "charitable trust", with about ten managing academic and government institutions controlling who can use the biobank's services and funnelling the profits back into the national health system. Western Australians will handle the studies and run trials for collaborators. No data, or access to raw data, will be sold.

Unappealing as that may sound to the pharmaceutical industry, the arrangement seems to work, says Eric Schadt, director of genetics at Rosetta Inpharmatics — a subsidiary of Merck — in Seattle, Washington. Schadt collaborated with deCODE to probe the genes that underlie obesity and says that the company was "extraordinarily cautious" about the protecting the privacy of Icelanders and only gave analysed results back to Rosetta. "I don't think you need your hands all over the data to get value out of them," says Schadt. If, like deCODE, Joondalup provides "high-value, scientifically solid results" that withstand scientific scrutiny, Schadt thinks that "most companies would be happy with that".

They seem to be. The US Healthcare IT provider Cerner and IBM are already partners, and Merck, GSK and AstraZeneca have all been eyeing Joondalup. Dan Burns, senior vice-president of pharmacogenetics at GSK speculates that the Joondalup study will place gene-disease associations in context: "There are plenty of examples of genes or genetic risks associated with diseases. But we don't know what that means, so the game is rapidly moving to function."

DeCODE led the way, says Burns, but it "got started around this belief that the goal was to identify the gene and I think the whole community has matured in its thinking. Finding a gene is a very important step along the way, but it is only a step."

> **"Finding a gene is a very important step along the way. But it is only a step."**
> — Dan Burns

## Power in the numbers

Joondalup, says Schadt, has the potential to be more powerful than deCODE. The key is in the in-depth phenotyping. "Why haven't we seen 15–20 highly replicated [genomewide association study] results for obesity like we have seen for diabetes? The reason, I think, is that it is an incredibly complex disease involving so many different parts of the system that the only way to really get a handle on that is to partition populations based on phenotype." The more phenotypes, the easier it is to stratify populations into subtypes and to get a handle on more complex diseases such as asthma or obesity.

Phase 4 clinical trials — post-marketing safety surveillance — will also be possible, says Palmer. "We know what drugs everyone has been on for the past 20 years, and we will have characterized them more carefully than any population has been characterized. So we can look for the subtle effects of two drugs interacting." That's going to appeal to pharmaceutical companies hoping to avoid debacles such as the one over Vioxx.

The greatest challenge that Palmer faces, says Troy Pickard, mayor of the City of Joondalup, is achieving the high participation rate. Low recruitment can stunt a biobank at its root. Lower-than-expected participation during the pilot phase caused concerns for the UK Biobank, although they say they are now on track to meet their goal by 2010. In Joondalup, Pickard says, the community is keen to participate and with good marketing Palmer should meet his goals.

"Lyle is a very motivating guy," says Hudson, "his drive is going to make this happen." With his three years worth of funding in hand, Palmer is anxious to begin the study. But, the project is in limbo until the participating institutes can agree on intellectual property and governance. "Whenever you get lawyers involved there are delays," says Palmer, with a note of weariness that seems at odds with his upbeat demeanour.

Palmer admits that every day, he wakes up and expects the whole project to fall apart, but somehow it continues. Colgan is also anxious for the study to begin. For a woman most definitely in control of her family's day-to-day schedule, health represents the only unknown. Colgan says that she knows which dietary and lifestyle choices are right for her family, but she's concerned about all the genetic factors that might be beyond her control. She says she'll do anything to ensure the best health for her children. As for all the testing? She says "the kids will probably like it". ■

**Bijal Trivedi is a freelance writer based in Washington DC.**

E. KENNEDY/NAVAL RES. LAB.

The HAARP facility includes 180 antennas.

closed them down in 2004, saying that they were no longer needed.

Another approach to ELF submarine communication was to take advantage of electrojets — currents of charged particles that flow through the ionosphere and could act as a virtual antennas, transmitting messages to submarines. Once this idea was proven experimentally[1] in the mid-1980s, physicist Dennis Papadopoulos, then of the Naval Research Laboratory in Washington, DC, began trying to drum up support for a new facility.

At the time the Pentagon was shutting down over-the-horizon radar sites that had been designed to detect Soviet bombers attacking the United States — including one in Gakona, an ideal location because it is underneath an electrojet. So Papadopoulos, who is now at the University of Maryland in College Park and has served as a scientific adviser for HAARP since the project's inception, argued for building an ionospheric heater there. The facility would help the Navy to study ELF waves, it would provide scientists with an ionospheric heater and it would guarantee continued life for the military site in Alaska, something that Alaskan Senator Ted Stevens, famous for steering congressional dollars to his home state, also liked. "That," says Papadopoulos, "was the genesis."

But even before construction began, people started to speculate about what the facility could be used for and why it was being built. In a news conference in 1990, Stevens talked about bringing energy from the aurora borealis "down to Earth so it could be used" to solve the world's energy crises, earning him the mockery of physicists. Others such as Nick Begich, the son of another Alaskan lawmaker, began claiming that HAARP was really intended as a missile defence weapon. According to Papadopoulos, these claims, although far-fetched, were based on a sliver of truth: Bernard Eastlund, a consultant to one of the firms building HAARP, had filed a series of patents making extraordinary claims that HAARP-like technology could be used as a defence shield by transforming natural gas into microwaves, which would knock out incoming Soviet missiles. The idea, jokingly dubbed the "killer shield", was even reviewed by the JASON defence advisory group, but was dismissed as "nonsense", according to Papadopoulos.

## From annihilation to defence

With the breakup of the Soviet Union, submarine communications no longer seemed as crucial, and HAARP needed a new *raison d'être*. Supporters proposed new tactics, such as studying ELF waves' ability to map out underground bunkers like those found in North Korea, a goal that quickly drew scepticism.

> "Scientific research to better understand Earth's ionosphere is a worthwhile endeavour." — Philip Coyle

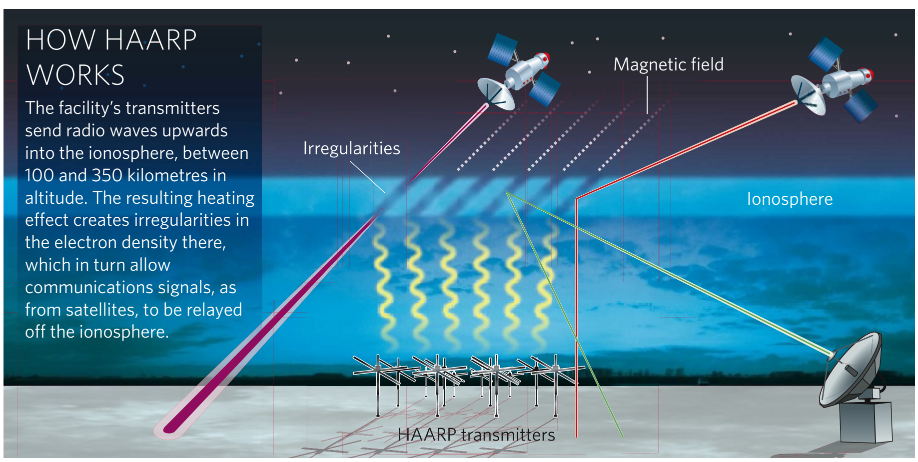
After the terrorist attacks of 9/11, however, the military found a new use for HAARP. In 2002, a panel headed by Anthony Tether, the director of DARPA, recommended that the facility be used to study ways to counter the effects of a high-altitude nuclear detonation, which would release energetic electrons that could cripple low-Earth satellites.

Electrons are produced naturally in this region when the solar wind, a stream of energetic particles flowing from the Sun, slams into the magnetic envelope that protects Earth. The planet has its own self-cleaning mechanism to rid itself of the particles: it eventually dumps them lower into the atmosphere through natural auroras and lightning. Scientists are now looking at whether they can accelerate this process by creating 'whistler' waves, which would kick the electrons into low enough altitudes — around 100 kilometres — where they would rain out naturally.

No one knows for sure whether it will work. "It is what we call a data-starved area — theory is ahead of actual observations," says Paul Kossey, HAARP's programme manager at the Air Force Research Laboratory at Hanscom Air Force Base, Massachusetts. Several experiments are being done to look at this possibility. Stanford University in Palo Alto, California, for example, is involved in the One Hop Experiment, which uses HAARP to inject very-low-frequency waves into the magnetosphere to create whistlers. The investigators use a buoy and ships in the South Pacific, where the waves fall back to Earth, to measure the presence of whistler waves[2].

Mitigating the radiation from an atmospheric nuclear detonation would require an entirely new facility, and the technology would be daunting. In 2006, a New Zealand-led group of scientists published a paper[3] arguing that any attempt to remediate radiation could lead to worldwide blackouts of high-frequency radio waves, disrupting communications and navigation. And some say that countering such high-altitude nuclear detonations is simply unrealistic. "I think scientific research to better understand Earth's ionosphere is a worthwhile endeavour," says Philip Coyle, a former associate director of the Livermore laboratory who served as the Pentagon's chief weapons tester during the administration of President Bill Clinton. But, he adds, they don't know how much energy they would need to flush the electrons, or how, ultimately, injecting this much energy would change the ionosphere.

In the meantime, there are plenty of straightforward science questions for HAARP to look into. The ionized part of the atmosphere has




HOW HAARP WORKS

The facility's transmitters send radio waves upwards into the ionosphere, between 100 and 350 kilometres in altitude. The resulting heating effect creates irregularities in the electron density there, which in turn allow communications signals, as from satellites, to be relayed off the ionosphere.

SOURCE: AFRL/ONR

long captivated researchers, going back to the days of Nikola Tesla, who dreamed of using it to send electricity around the world. In 1933, scientists found that changing the electron density in the ionosphere could alter the propagation of radio signals[4]. That discovery eventually led to the development of ionospheric heaters to study these and other effects.

## Bells and whistles

Radiation from solar flares is one area of interest. "These things are really important because it is the radiation coming off the Sun that is the main cause of satellite failure or potential death in human space exploration," says Michael Kosch, the deputy head of the communication systems department at Lancaster University, UK. Other areas include looking at the processes that cause an aurora — when electrons in the magnetosphere collide with the uncharged particles of the atmosphere, creating the optical emissions often seen as brilliantly coloured lights in the night sky. One of HAARP's most cited accomplishments is the creation of the first artificial aurora visible to the naked eye[5]. On zapping the ionosphere, HAARP created a green aurora between 100 and 150 kilometres high — in the middle of a natural aurora. "That was something you couldn't predict," says Michael Kelley, a physicist at Cornell University in Ithaca, New York, who has been involved with HAARP.

Other ionospheric heaters around the world include a lower-power US facility in Arecibo, Puerto Rico, which has been offline since a flood several years ago (although plans are under way to refurbish it), and one in the Russian city of Vasilsursk, which has struggled with funding issues. HAARP's closest peer is a powerful ionospheric heater at the European Incoherent Scatter (EISCAT) Scientific Association in northern Scandinavia. EISCAT's heater has cost roughly $24 million to build and operate to date, and was the first to create


an artificial aurora, even before HAARP.

HAARP, though, has the highest power as well as the most advanced optics and diagnostic equipment. But most of all, its phased-array radar means that the signals can be steered and controlled digitally. It can also create multiple beams, which can be shaped, or changed instantaneously to sweep north, south, east and west. "I think the main thing that makes it unique is that it has a much wider frequency operating range," adds Kosch, who has also worked extensively at EISCAT. HAARP operates between 2.8 and 10 megahertz, whereas EISCAT operates between 3.9 and 8 megahertz. "It can operate in a much lower frequency range than the one we can use here in Europe," Kosch says.

> **"HAARP can operate in a much lower frequency range than the one we can use here in Europe."**
> — Michael Kosch

As HAARP was only finished in 2007, scientists and Pentagon officials involved in the project concede that management issues, such as allocating time at the facility, are still in the formative stages. In fact, one of the most recent HAARP experiments is something that's not likely to show up in the scientific literature at all: an experiment done in January that involved sending radio waves to the Moon and then having amateur radio enthusiasts and a receiving antenna in New Mexico measure the reflected signals. But Papadopoulos says that the experiment was more for the amateur radio community than for scientists.

At the moment, time at the facility is divided between researcher-directed work, which takes place during 'campaigns' of two to three weeks, and military needs. "It's a fairly complicated situation in which we support new researchers, and new people, by getting them involved in the campaigns, which is relatively cheap," says Kossey. "Then of course we also fund [military] proposals and contracts that come in under broad agency announcements, in which researchers propose research that is of interest to the various organizations."

And even though HAARP is a military-owned facility, academics say that access has not been a problem. Umran Inan, the lead scientist for the Stanford work, says that Stanford has been one of the most frequent users, with numerous graduate students and foreign scientists working at the site. "Obviously, there are security arrangements, because it's a US Department of Defense facility," says Kosch. "I'm a foreigner — escort required — but I am already so familiar to the people there, and so familiar with the facility, that it's not really a major problem."

HAARP's evolution may not have been straightforward, but it is, in the minds of many scientists who work there, a success. "HAARP has been a boon to science in this area, and I think the managers that run HAARP, from the very beginning, have involved the community," says Inan. So unlike many other Department of Defense facilities that are built before there is a clear rationale, "in this case the community was involved from the very beginning, so the properties of the facilities were all defined with the involvement of the community. Now, I think it's a thriving success," he says.

As for HAARP's original legacy, as an antenna to send signals to submarines, that era has come and gone with the end of the cold war. "The communications for submarines is not as important any more," says Papadopoulos. "There are," he acknowledges, "no submarines from the other side." ■

**Sharon Weinberger is a freelance writer in Washington DC.**

1. Barr, R., Rietveld, M. T., Kopka, H., Stubbe, P. & Nielsen, E. *Nature* **317,** 155–157 (1985).
2. Inan, U. S. *et al. Geophys. Res. Lett.* **31,** L24805 (2004).
3. Rodger, C. J. *et al. Ann. Geophys.* **24,** 2025–2041 (2006).
4. Tellegen, B. D. H. *Nature* **131,** 840 (1933).
5. Pederson, T. R. & Gerken, E. A. *Nature* **433,** 498–500 (2005).

B. MARTINSON/AP

**Scientists want to better understand the processes involved in creating auroras.**

## Media right to report small value of antidepressants

SIR — Your Editorial 'No more scavenger hunts' (*Nature* **452,** 1; 2008) indicates that the press reaction to a recent meta-analysis of antidepressant efficacy was exaggerated — and that it was "not the media's finest hour". But I do not find their reaction surprising.

To one who has followed the scientific literature over the years, the fact that antidepressants' superiority over placebo is at best modest may come as no surprise: in fact, this is apparent in most published clinical trials. But the portrayal of antidepressants as advertised by pharmaceutical companies, ratified by medical professionals, propagated by mass media and absorbed by the general public during the past two decades has been very different. They are promoted as highly efficacious drugs (not true, when considering effect magnitude) that correct the 'chemical imbalances' underlying depression (not true, when considering how little we know about its pathophysiology) and improve depressive symptoms better and faster than psychotherapy (not true, when considering most clinical trials that compare the two). This publicity prompted an even more misguided rejection of antidepressant therapy by some of its opponents, with claims that the drugs would 'dope' people into happiness and prevent them from working out their problems.

Publicizing the modest effects of antidepressants goes some way towards countering the much more exaggerated positive hype that preceded it. Although it might not have been the media's finest hour, on this subject it was still finer than most.

**Olavo B. Amaral**
Departamento de Bioquímica, ICBS, Universidade Federal do Rio Grande do Sul, R. Ramiro Barcelos 2600 (anexo), 90035-003 Porto Alegre, RS, Brazil

## Tropical farmers need productive alternatives

SIR — Your Editorial 'Markets can save forests' (*Nature* **452,** 127–128; 2008) proposes integrating deforestation into the international carbon market, but it is unlikely that money alone can solve the problem.

The process of tropical deforestation presents a dilemma: enormous economic value (more than US$2,000 a year per hectare) is lost in favour of small private benefits (often less than US$100 a year per hectare). In a true market, an increased, scarcity-signalling price should provide an incentive to boost the supply of the scarce commodity. So far, this has not worked in the case of tropical forests. Real financial flows received for the provision of ecosystem services are vanishingly small, if they exist at all — only a few beneficiaries have been convinced to pay for environmental services. This illustrates the essentially theoretical character (at least, up till now) of economic values generated by ecosystem services.

'Payments for ecosystem services' are supposed to save tropical forests. But even if it were possible to mobilize substantial payments, tropical farmers wouldn't be prepared to stand by and twiddle their thumbs while receiving them. They need a field of activity, so sustainable land-use concepts should address the social environment and needs of people as well. These concepts could be linked and possibly financed by such payments.

If the people who use tropical lands are disregarded, we don't believe that payments for ecosystem services will solve the problem of disappearing tropical forests. What is needed is an economic system that keeps people gainfully employed in an activity that is ultimately productive, not destructive.

**Thomas Knoke*, Bernd Stimm†, Michael Weber†**
*Institute of Forest Management, Technische Universität München
†Institute of Silviculture, Technische Universität München, Am Hochanger 13, 85354 Freising, Germany

## Researchers should explain why they use animals

SIR — Your News story 'Animal-rights activists invade Europe (*Nature* **451,** 1034–1035; 2008) highlights the need for medical researchers to do more to communicate to the public the reasons why they need to use animals in their research and what this involves. All too often, there is a tendency to wait until extremism becomes intolerable before taking steps to counter it (see *Nature* **452,** 282; 2008). The little information about animal research available to the public is frequently oversimplified and tends to be over-reliant on the perceived authority of the author. The scientific literature usually requires subscription to access it and scientific training to understand it. This leaves information gaps through which antivivisectionist groups can push their propaganda.

Organizations such as the Research Defence Society (www.rds-online.org.uk) do much to address this deficit, but have limited resources and cannot be expected to counter the animal-rights campaigners alone. Anyone who is wondering why somebody doesn't debunk misleading claims made about them or their colleagues should consider the possibility that they are that 'somebody'. Even those who are not prepared to go public can always provide detailed explanations of their work and that of others in the field to scientific advocacy campaigns.

A fact your report didn't mention is that the new biomedical laboratory in Oxford — which, by the way, will house mostly rodents and very few monkeys — has been built. In a campaign that complemented the efforts of the police and government, Pro-Test were able to counter the animal-rights group Speak ('The voice for the animals') by capitalizing on the overwhelming support for the new laboratory among Oxford students and local politicians.

Extremism can be defeated, but only if scientists stand up and expose the myths and distortions that fuel it.

**Name and address supplied**

## Truth about a plant with many names

SIR — In the portrayal in Books & Arts of the University History Museum in Pavia, northern Italy (*Nature* **451,** 526, 2008), the naturalist Giovanni Antonio Scopoli — one of Pavia's many famous professors — is linked with the drug scopolamine. But although the two are connected, the compound was in fact named after the plant from which it is derived, *Scopolia carniolica*.

Indeed, Scopoli (1723–1788) was also a physician. He worked for a time at the huge 500-year-old mercury mine in Idrija in Carniola, now part of Slovenia, and recorded the adverse effects of mercury in miners. In this mountainous region, he studied the local plants and published his findings in the famous *Flora Carniolica* in 1760 and 1772, and corresponded in Latin with the Swedish naturalist Carl Linnaeus.

The scopolamine-containing plant was first described in 1569 by Pietro Andrea Mattioli and was identified by Scopoli in the forests around Idrija as *Lithophila*. The Vienna court botanist Nikolaus Joseph von Jacquin in 1764 changed the name to *Scopola carniolica*, in honour of the great naturalist. However, Scopoli avoided using the name, as 'scopola' is an insult in Venetian dialect, indicating a slap in the face. Linnaeus changed the name to *Hyoscyamus scopolia* in 1767, and in 1790 the genus name *Scopolia* was adopted.

About 100 years later, Ernst A. Schmidt at the University of Marburg isolated an alkaloid constituent from *Scopolia carniolica*'s dried rhizome, naming the drug scopolamine. This sedative has also been used by ophthalmologists and as a 'truth serum' during the Cold War.

**Marko Kreft, Robert Zorec**
Laboratory of Neuroendocrinology-Molecular Cell Physiology, Institute of Pathophysiology, Faculty of Medicine, University of Ljubljana, Zaloška 4, 1000 Ljubljana, Slovenia, and Celica Biomedical Center, Tehnološki Park 24, 1000 Ljubljana, Slovenia

COMMENTARY

# Europe's research system must change

Science funding in the European Union needs to be revised to better serve economic, social and environmental goals, **Luke Georghiou** argues.

For researchers in Europe, the Framework Programmes have become a familiar funding source for projects and mobility. They account, however, for only about 5% of total public-research money[1]. Since 2000, the European Commission has sought to extend cooperation beyond its own programmes into the remaining 95% of public research, which is funded by national governments. It is doing so under the concept of the European Research Area (ERA) — that we at the Commission's ERA Rationales Expert Group have advised[2] — using measures such as improving cross-border access to infrastructures, coordinating national research programmes and facilitating researcher mobility.

During the past year a consultative green paper has set out ideas for a relaunch of the ERA[3]. This followed some innovation in the Seventh Framework Programme, which included the foundation of the European Research Council (a funding body for investigator-driven research allocated through Europe-wide competition) and the industry-led Joint Technology Initiatives — such as the Innovative Medicines initiative for pharmaceuticals and the Clean Sky initiative for green aircraft manufacture and operation — which can be worth up to €1 billion (US$1.5 billion) each. In essence Brussels has proposed 'more of the same and better' to make the public-research system function more effectively.

These measures do not get to the core of why we need the ERA. They do not reach out beyond the research community to appeal to politicians, business or the general public. There is the chance of a major transfer of resources from agricultural subsidies to research and innovation with a new budget settlement and financial framework due for the European Union (EU). Having as the flagship policy one that focuses on remedies for perceived failings in the research system lacks ambition commensurate with this broader vision.

## Radical approach

We need a shift in thinking from deficit to opportunity. We must make a convincing case for increased investment in research by both the European Union and national governments. This will help us attain Europe's economic, social and environmental goals.

Making that case requires a radically new approach to European research. Three key areas of action are needed. First, Europe's research system must respond to a series of 'grand challenges'. Second, Europe must become more research-friendly. This requires major reforms for many types of institution and the policies that support them. It also means that the ERA should extend to private as well as public research. Third, Europe's strategic and applied research must be re-orientated at a pan-European level to support the full range of policies that member states have agreed. This involves the Framework Programme and national programmes — coordinated through ERA-NET schemes and other instruments — engaging much more effectively with policy needs in areas such as the environment, transport, energy, agriculture and health.

"We should engage research with the problems that society recognizes as central."

Basic research has an important place, and promoting excellence through the European Research Council, and building research capacity through mobility programmes are worthy goals that deserve more support. But the bulk of the Framework Programme budget is in strategic and applied research so we need to think more deeply about its aims and future direction. Officially the budget is there to support European competitiveness and public policies. Despite some significant achievements, successive evaluations strain to provide any overall picture of impact[4]. Existing structures have in general failed to provide the kind of linkage that could allow research to efficiently support economic and social priorities. National delegates return from European meetings with their performance indicator being the budget share their nationals obtain — '*juste retour*' — rather than the benefit their country sees from that money.

## Grand challenges

Historically Europe has been at its best when dealing with large projects based on public–private partnerships. These include: the development of the Global System for Mobile Communications — better known as GSM — as the standard for mobile telephony; the emergence of Airbus as a global player; the ubiquitous use of nuclear energy in France; economic leadership in wind energy; and scientific leadership at CERN, Europe's particle-physics laboratory near Geneva. Similarly, in the United States and Japan many world-shaping innovations have emerged from grand challenges and coordinated efforts, such as the atom bomb, spaceflight, semiconductors and the Internet.

Such projects created the conditions for

BORDAS/SIPA/NEWSCOM

Coordinated effort: European premiers Angela Merkel and Nicolas Sarkozy confer at the Airbus factory.

entrepreneurs and individual scientists or engineers to seize opportunities. Governments provided funding and constructed markets through effective regulation and procurement. Programmes may now need to be more agile to keep pace with rapidly moving fields; nonetheless the lessons of scale, vision and commitment are clear.

The grand challenges will probably require a more direct political appeal to get going. They should engage research with the problems that society recognizes as central, such as climate change, food and energy security and the ageing of western society. For these, the initiative will have to come from governments rather than business, although many business opportunities will emerge as initiatives unfold.

European governments cannot rely solely on the Framework Programme budget; they will have to coordinate their national budgets by accepting leadership of sub-sections of larger programmes, and by funding the participation of their own nationals at the very least. The challenges must be of a clear trans-national nature and require a minimum level of effort that cannot be achieved by nations acting alone. They must also be feasible: there must be a base of research and industrial capability to build on and a viable implementation path. And of course research must be a necessary and important part of the solution.

## Right condition, right coalition

An exemplary response to a grand challenge may have recently emerged. Achieving a low-carbon future is arguably the single greatest test facing us all. The recently announced European Strategic Energy Technology Plan (SET-Plan) describes itself as a "far-reaching jigsaw of policies and measures". These include binding targets for 2020: a 20% reduction in greenhouse-gas emissions; 20% of renewable energy sources in the EU energy mix; reducing primary energy use by 20%; carbon pricing; a competitive Internal Energy Market and an international energy policy. Central to this plan is the need to accelerate the development of cost-effective low-carbon technologies.

The challenge here is great. Energy research has stagnated for decades, with total EU public spending falling to a quarter of 1980 levels in real terms and substantial declines in the private sector as well[5]. Energy innovation is particularly difficult, involving very large investments with long lead times and lock-in to existing infrastructures. New technologies face social acceptance issues and often begin by being more expensive than the sources they are intended to replace.

On the other hand, addressing climate change offers huge opportunities to develop business and employment. With similar investment decisions in other major economies there is a strong competitiveness rationale. The SET-Plan has also developed a series of key technology goals for 2050. These include second-generation sustainable biofuels, $CO_2$ capture, and large-scale commercialization of renewable energy, energy conversion and efficiency. A high-level multi-government steering group will oversee the SET-Plan implementation, engaging with all stakeholders. Among other actions a series of industrial projects resembling the Joint Technology Initiatives will be launched. The hardest part will be securing the necessary finance.

We need ways to identify and mobilize similar, necessary coalitions of interest in other areas. Arguably, the SET-Plan has arisen from a once-in-a-lifetime coincidence of practical urgency, political will and technological opportunity. With grand challenges potentially costing €5 billion to €10 billion each, it is hard to see Europe affording more than three or four at a time even with an increased budget.

To pinpoint these ways, a new kind of political process is needed that combines top–down and bottom–up approaches. A bottom–up phase would encourage stakeholders to form 'platforms' to develop potential responses to challenges. Like the present technology platforms, these would have a wider base of participation: science; business users and suppliers; government policy-makers, regulators and purchasers; and where relevant, non-governmental organizations and consumer groups. These would use targeted foresight to bring together socioeconomic demand and the potential of innovation and act as both incubator and lobby. The aim is not to follow the now discredited idea of picking winners among firms or even technologies (e.g. joint programmes for high-definition television that have now been abandoned), rather to create a competitive and supportive environment in which winning solutions emerge. The top–down element will require, at the highest political level, a capacity to find resources very quickly when a viable strategy has emerged. The core budget would come from a fund deployed until exhausted[6].

"Europe will be a much more effective research partner if it can speak with one voice."

## Policy-focused

There is currently only a very general communication between the research carried out at European level and the European-level policy and regulation setting. This is true both for the Framework Programme and for the ERA-NETS that have begun to link national funding bodies. Indeed it is a problem that besets much national applied research. New kinds of coordination are needed to link users and sponsors of research, at European and at national levels. This does not mean a crude customer–contractor relationship that often causes research to degenerate into consultancy. It does mean that regulating bodies, such as the sectoral Directorates-General, will need a greater voice in establishing the research agenda and greater scientific capability to do so.

Neither the grand challenges nor the policy-focused research can be achieved through the present research system. This is where the new ecology comes in. It must consist of reformed actors and better linkages between them to configure research around these interdisciplinary challenges[7]. The long list of reforms that are overdue includes: giving greater strategic space and autonomy to universities; more trans-national peer review to raise quality levels; developing a true European market for applied research services (cross-border trade in applied research accounts for a negligible share of a market worth billions); and creating a market friendly to innovation through smart regulation and public procurement[8].

Europe has to let go of structures and approaches that have dominated its research funding for decades. First to go should be Framework Programmes that are divided into large numbers of small, very loosely connected projects defined years ahead by 'work plans' with no clear provenance. These instruments may satisfy the clientele they fund, but they are almost impossible to direct towards real problems. The funding breakdown needs to be tied to the big- and the medium-level challenges that policy dictates. Enough flexibility must be retained to respond to shifts in demand and to new scientific and technological opportunities. Researchers worried about losing scarce funding should recognize that those who can adapt stand to receive slices of potentially a much larger cake.

Europe should start the process of reform now. The Commission has a responsibility to take the lead and planning of the Eighth Framework Programme, due to start in 2013, is already under way. The grand challenges will not wait until then and member states, businesses and the scientific community must each play their part. The first challenge is one of leadership.

Those outside Europe who might see this discussion as parochial should consider this: Europe will be a much more effective partner for the United States, Asia and others if it can speak with one voice, take the initiative and contribute a genuine critical mass to solutions to global problems. ■

Luke Georghiou is in the Manchester Institute of Innovation Research at Manchester Business School, University of Manchester, Manchester, M15 6PB, UK. He also chairs the European Commission's ERA Rationales Expert Group.

1. Mustar, P. & Esterle, L. *Key Figures on Science and Technology* (Observatoire des Sciences et des Techniques, Paris, 2006).
2. http://ec.europa.eu/research/era/progress-on-debate/expert-groups-analyses_en.html
3. http://ec.europa.eu/research/era/pdf/era-greenpaper_en.pdf
4. European Court of Auditors *Evaluating the EU Research and Technological Development (RTD) Framework Programmes — could the Commission's approach be improved?* Special Report No 9/2007.
5. http://ec.europa.eu/energy/res/setplan/doc/com_2007/com_2007_0723_en.pdf
6. Larédo, P. *Sci. Pub. Pol.* **30,** 4–12 (2003).
7. Coombs, R. & Georghiou, L. *Science* **296,** 471 (2002).
8. http://ec.europa.eu/invest-in-research/pdf/download_en/aho_report.pdf

nature

BOOKS & ARTS

# Orange revolution

**Stories of seventeeth-century scientists and aristocrats show how Dutch ingenuity benefited England.**

**Going Dutch: How England Plundered Holland's Glory**
by Lisa Jardine
HarperPress: 2008. 400 pp. $35, £25

**Harold Cook**

History is often told from a national perspective, but big ideas usually have cross-border entanglements. Lisa Jardine's carefully crafted and highly readable book describes how people and concepts from the Netherlands percolated English high culture in the seventeenth century, influencing early science. *Going Dutch* may unsettle those raised on the parochial view of the English as driving their own independent destiny.

Historian Jardine begins with the Dutch invasion of England known as the Glorious Revolution of 1688–89. No blood was shed, yet England was subjected to a massive *coup d'etat* at the point of a foreign prince's sword. The head of the Dutch army, Prince William of Orange (also the nephew of England's King James II), gathered a fleet of more than 500 ships to convey his battle-hardened troops across the water, an operation the size of which was not repeated until D-Day in 1944.

Marching on London, the prince was greeted by cheering crowds. Meanwhile, James II's army withdrew rather than offering battle. A cabal of Protestant lords provided political cover by inviting William to take over the English government. The imprisoned James II was allowed to escape to France, while a hastily convened Parliament pronounced William and his wife Mary (daughter of James II) as joint sovereigns, giving legitimacy to the new regime. But William's Dutch guard garrisoned an occupied London for years afterwards, just to make sure.

Why was this quiet coup seen as importing a king rather than suffering a conquest? Jardine argues that Dutch victory was subverted by English opportunism. By 'going Dutch' and adopting the commercial and administrative methods of their new masters, the English quickly gained the upper hand, replacing the Netherlands as the major international power. Jardine suggests that a common cause was possible between the sometime enemies because the countries were culturally close.

Providing a family history of the English and their Dutch first cousins, Jardine explores personal networks between influential characters. The main vehicle is the Huygens family, including elder statesman Constantijn and two of his sons Constantijn and Christiaan, the notable scientist. Constantijn Huygens junior accompanied William during the invasion of England, whereas Constantijn senior, a long-lived Anglophile, had served as the principal secretary to the house of Orange for many decades before.

Christiaan Huygens's scientific reputation was boosted by English boasts.

COLLECTION HAGUE HISTORICAL MUS., THE HAGUE, THE NETHERLANDS

The two chapters concerning the scientific work of Christiaan Huygens and Robert Hooke are the most original. Jardine emphasizes the exchanges between the virtuosi of England and the Netherlands that amounted to an international scientific forum, even through the period of the Second Anglo–Dutch War of 1665–67. Two case studies form the book's core: the debates about the accuracy of Huygens's pendulum clock for finding longitude in the 1660s and his balance-spring watch of the mid-1670s, and the discussions around Robert Hooke's famous book *Micrographia*, reproducing microscopic biological observations in exquisite detail.

The careful reconstruction of events surrounding the adaptation of Huygens's clock shows how much he depended on the innovations and experiments of his English friends. Jardine discovered new evidence in Samuel Pepys's papers about how sea-trial reports to the Royal Society regarding the pendulum clock were exaggerated by Admiral Robert Holmes. In doing so, Holmes, who helped to start the Second Anglo–Dutch War, ironically gave the Dutchman Huygens a claim to priority that obscured the contributions of the English.

Showing the further intertwining of Anglo–Dutch intellectual networks with those of France, Jardine demonstrates how in the mid-1660s, Henry Oldenburg, the first secretary of the Royal Society, tried to raise doubts about Hooke's *Micrographia* internationally. This helped to undermine Hooke's reputation, contributing to the later failures of Huygens and Isaac Newton to acknowledge his contributions. Jardine remains one of Hooke's chief advocates, placing him and Huygens on the international stage.

Jardine's circles move outwards beyond the Huygens family and science to links between the house of Orange and the Stuarts, and to the lives of English royalists in exile in the Low Countries during Cromwell's government. Although this creates the impression that both countries were tied by conversations and intermarriages among the great and the good (genealogical tables are supplied in the book's appendix) rather than by the connections of ordinary people, Jardine's strategy lets her highlight many topics without resorting to generalizations. She addresses fine art (mostly painting), music, gardening and botany, science and commerce, accompanied by colour illustrations.

*Going Dutch* is richly evocative. One feels present at a masque in The Hague sponsored by

nobles of the Winter Queen's court, with dancing until 4 a.m., or walking through the estate garden of the elder Huygens, or accompanying his third son Lodewijk through Somerset House in London after the execution of Charles I to view the impressive royal art collection.

This fascinating book is an excellent introduction to seventeenth-century Anglo–Dutch relations. Jardine does not try to summarize the whole field. She avoids, for instance, examining the Anglo–Dutch wars of the period or the bitter rivalries abroad, gives much more space to royalists than republicans, scarcely deals with religion, and treats the formative period of English philosopher John Locke's exile in the Netherlands very lightly. She does not develop fully an account of how the Dutch coup launched the Bank of England, nor how it affected Scotland and Ireland. These subjects are left to other works, many of them cited in her bibliography. Jardine presents the view from England more than that from across the North Sea, and her subtitle is perhaps more relevant to the eighteenth century than the seventeenth. Yet by exploring pertinent examples, *Going Dutch* demonstrates that personal connections helped to shape the cultures of both countries. ■

Harold Cook is professor at and director of The Wellcome Trust Centre for the History of Medicine at University College London, London NW1 2BE, UK, and author of *Matters of Exchange: Commerce, Medicine and Science in the Dutch Golden Age*.

# Biased brains, messy memories

**Kluge: The Haphazard Construction of the Human Mind**
by Gary Marcus
Houghton Mifflin: 2008. 224 pp. $24

**A Portrait of the Brain**
by Adam Zeman
Yale University Press: 2008. 256 pp. $27.50

**Sandra Aamodt**

Public interest in the brain seems to be insatiable, judging from the many popular books about the topic that have been published in the past few years. Highlighting the diversity of this expanding genre, two books aimed at general readers provide views of brain science in very different styles.

In *Kluge*, psychologist Gary Marcus presents a lively tour of the shortcomings of human minds and concludes that evolution has left us with something of a mess. In an argument reminiscent of David Linden's *The Accidental Mind*, Marcus makes his case by describing cognitive difficulties, including false beliefs, linguistic ambiguity, impulsiveness and mental illness.

The blame, he asserts, rests with our imperfect memory, "arguably the mind's original sin". Perhaps we would reason more effectively if the brain could store and retrieve data as accurately and as simply as a computer. Instead we must contend with a limited system. Brains locate memories by matching them to the current context rather than having unbiased access to all of our experiences. This contextual dependence makes it hard during an argument, for example, to recall how often our spouse does the housework, because thinking of one failure inclines our brains to remember similar situations rather than contrary examples.

Many of these problems result from conflicts between the brain's two basic styles of thinking. The reflexive system, having evolved earlier, controls most behaviour. It is fast and can accurately assess statistics — such as the likelihood of finding food in certain locations — but is prone to overgeneralization and snap judgements. The deliberative system, by contrast, is slow, effortful and logical, at least intermittently. The reflexive system readily overrides the deliberative system, especially when we are tired or rushed. Marcus believes our lives would be improved if we engaged the deliberative system more often, although he acknowledges that "it often settles for reasoning that is less than ideal".

Evolutionary psychology has tempted many scientists to indulge in just-so stories, as Marcus notes; asserting that our brains are poorly engineered is an equally risky business. Computer memories are more factual than those of humans, but computers lag far behind the reflexive system on other problems, such as distinguishing cats from dogs. Whether any biological system (or a computer) could combine both sets of virtues remains an open question.

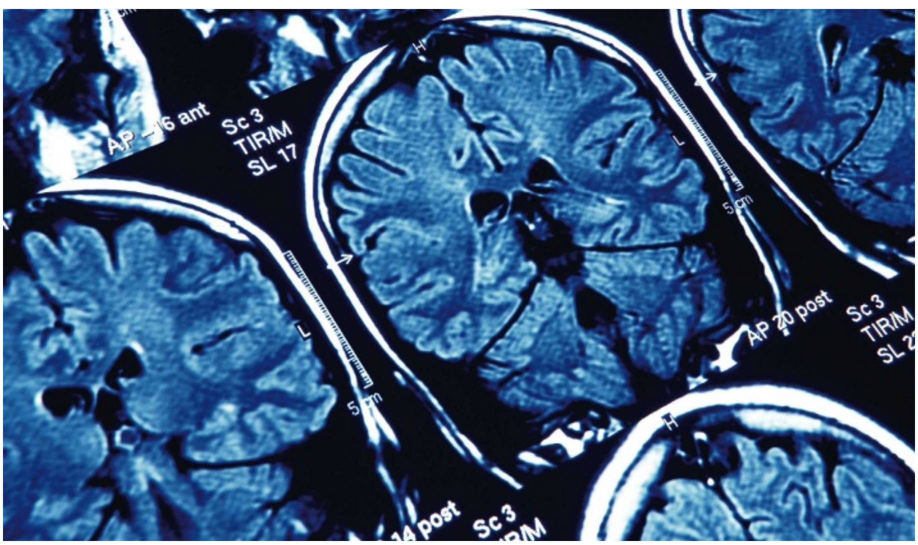
Marcus closes with experimentally verified ways for people to deliberate more effectively, a welcome change from the usual self-help prescriptions. For example, he suggests imagining that your decisions might be checked by someone else. Although few biologists will need to be convinced of the evolutionary arguments in this book, it remains intriguing to contemplate which aspects of our minds could be improved and how we might compensate for those weaknesses.

*A Portrait of the Brain* is constructed with equal care, but in a very different style. Neurologist Adam Zeman takes on an ambitious project: explaining brain function from atoms to neural networks to, unexpectedly, the soul. Clearly an Oliver Sacks fan, Zeman weaves case studies of patients together with basic science, history, etymology, classical literature and art to produce an erudite discourse on brain components.

Describing the many aspects of neurons, the fifth chapter explains, for instance, how early neuroanatomists determined that brains are composed of cells, tells the story of a bus driver with religious feelings caused by epileptic seizures who ultimately became a priest, and is laced with asides about scientific rivalry, brain development, the hobbies of neuroscientist Santiago Ramón y Cajal and how action potentials are generated. The science is clear, and the stories of patients are suspenseful and gripping. Yet in some places, the relationship between topics feels a bit forced.

In the final chapter, Zeman grapples with consciousness. He outlines how brains that are predisposed to tell stories and that attribute actions to agents rather than chance might lead us to believe in an immortal soul. His own view is that this is "no more than a wonderful fiction". (Marcus makes the same point less gently.) Zeman struggles with science's failure to find an emotionally satisfying replacement story, conceding that such questions may be more in the realm of art than science.

What is the impetus for this deluge of brain




PASIEKA/SCIENCE PHOTO LIBRARY

**Powers of persuasion: human brain images have piqued public interest in neuroscience and psychology.**

books? Applications outside of science may explain the popularity of some topics, such as the potential for behavioural economics to illuminate the decisions of investors. Brain-scan images themselves may also have amplified interest in neuroscience. Both the public and the media are drawn to the powerful and persuasive visual message of such images. Fundamentally, however, people remain interested in neuroscience and psychology because understanding our brains helps us to understand ourselves. ■

Sandra Aamodt is former editor of *Nature Neuroscience* and author of *Welcome to Your Brain: Why You Lose Your Car Keys But Never Forget How to Drive and Other Puzzles of Everyday Life*.

# Technological twist on taxonomy

**Systematics as Cyberscience: Computers, Change, and Continuity in Science**
by Christine Hine
MIT Press: 2008. 320 pp. $35.00, £22.95

**Kevin Kelly**

Taxonomy, the science of identification and classification of new species, has been one of the slowest disciplines to adopt computers. When most other scientists routinely use these number crunchers to detect patterns within large sets of data, why have taxonomists only recently started to use them?

The reasons are many. Foremost has been the subtle variation among closely related species, which makes quantification of their traits difficult. No computer program can outdo the highly refined judgements of a taxonomic expert who can classify from nuanced alterations even the smallest organism. Consequently, new species are identified and described in a manner that would have been familiar to Charles Darwin 150 years ago.

Second, much taxonomic information has been, and remains, parochial. The expertise required for classifying fly parasites has little in common with that for fungal species or whales. Taxonomic information occupies niches — niche being the exact biological term for these narrow confines. Specialized niches of information with their own protocols challenge computerization.

Third, the low priority given to taxonomy has meant it is perennially underfunded. High-powered computation and software come low on the list after the meagre needs of traditional taxonomy are (barely) met.

Despite these hurdles, the related field of systematics (exploring relationships between organisms over time) is rapidly transforming itself as computation becomes integral. In *Systematics as Cyberscience*, sociologist Christine Hine investigates the effects of computers and communication technology on the taxonomic community.

Hine's subjects are primarily the practising taxonomists and systematists in the United Kingdom, but this community is representative of that in the rest of the world. Her concern is the sociology of change, how these tradition-bound disciplines imagine, place and incorporate computers into their identity.


The Encyclopedia of Life will log facts for all species on Earth.

D. BACKLUND/WILD PHOTOS PHOTOGRAPHY; J. SULLIVAN/WILD HERPS; IMAGES COURTESY OF EOL

For instance, she notices that "the Web and the Internet are seen … as evocative objects which bring a certain glamour and an image of modernity and accessibility to systematics". Computer technology is not only an aid to getting work done, and to getting new kinds of work done, but also to securing the respect and funding needed to do any work at all. This strategy plays into a misperception that you are not doing real science unless you are doing it on computers.

As a consequence of digitalization, classical taxonomy is put under different strains as it adjusts to its new tools. Hine notes that: "Traditionally, systematics has worked on long time scales, with taxonomists focusing on grand life works, and major flora projects taking maybe 30 years to complete." Now, by contrast, taxonomy projects based on Internet databases are offered up in minutes, and are never viewed as finished. Instead they remain in the 'perpetual beta' state of most digital content.

Cheap, ubiquitous genetic sequencing is further accelerating taxonomy's rush into computation. Classifications based on gene sequences are hugely disruptive to the established order of doing things. The focus of *Systematics as Cyberscience* is not rapid genetic taxonomy, which could fill its own book, but these recent advances only heighten the identity shift now under way in taxonomy.

As taxonomy evolves into the cyber age, it is reappearing in the public eye. In February 2008, systematists launched the first digital version of the Encyclopedia of Life (www.eol.org), wherein many of the approximately 1.8 million known species on Earth will be represented, each on its own web page. The fact that most of these pages are still blank does not dilute the mythical power of this embryonic encyclopaedia. We can see its promise in the 30,000 populated pages now in operation. Each species page includes photographs of the organism in its various life stages, the original taxonomic description, updatable geographical maps of its range, classification charts, links to ecological niches, popular knowledge of the organism, and many other attributes needed to understand its biology.

The most important aspect of this emerging cyber library is the way in which it facilitates the completion of the great taxonomic imperative: to identify all of the living species on Earth. There is doubt as to how long this will take, but it will surely happen only with the help of vast communication webs and computer technology.

The future of systematics will involve advances in computational photography, innovations in creating identification keys, pattern-recognition software, expert-knowledge systems and, eventually, genetic identification gadgets. Soon, the practice of taxonomy will be unthinkable without the Internet. Yet, as Hine notes, at the moment "systematics has been able to portray itself as fundamentally unchanged by its experience with these technologies, stressing instead the way it is pursuing its established goals and preserving its heritage."

The long sweep of evolution breeds in those who try to untangle it a great appreciation of the winnowing process of time. Taxonomists are wary of tossing away any highly evolved method of knowledge that has proven itself. It is remarkable then to see the speed at which these refined practices are being reimagined for this digital century.

*Systematics as Cyberscience* reminds us that change generates strain, conflicting views, fear of the new and concern for integrity. But we cannot stop this shift that is being played out by experts in their respective museums and natural history collections. ■

Kevin Kelly is senior maverick at *Wired* magazine and chair of the All Species Foundation.

NATURE|Vol 452|24 April 2008

# Hidden treasures: the Cajal collection in Madrid

The perceptive drawings, paintings, photographs and slides of Spain's neuroanatomy pioneer record a tale of ambition and rivalry, reports **Alison Abbott**.

In 1906, the two Nobel laureates in medicine brought their hostilities to the prize ceremony. Italy's introverted Camillo Golgi used his speech in front of the King of Sweden to launch a vicious attack on Spain's Santiago Ramón y Cajal and his (correct) theory that the brain is composed of billions of individual cells, or neurons. Cajal spoke only of his scientific observations, but his presentation rang a deliberate death knell for Golgi's alternative theory that the brain was composed of a continuous web of interconnected cells.

The Spaniard was fire to the Italian's ice. Outgoing and passionate, Cajal was a latecomer to neuroanatomy. He may have been a teenage rebel, but he bent to his domineering father's demands to abandon all notions of being an artist, and studied medicine instead. After his training, Cajal joined the military to fight for Spain in the Ten Years War in Cuba, and came back with malaria, which then became complicated by a tuberculosis infection. Not surprisingly, his first scientific ambition after joining the staff of the University of Saragossa, Spain, in 1875 was to understand and treat such infectious diseases. With the money saved from his military service in Cuba, he bought his first microscope.

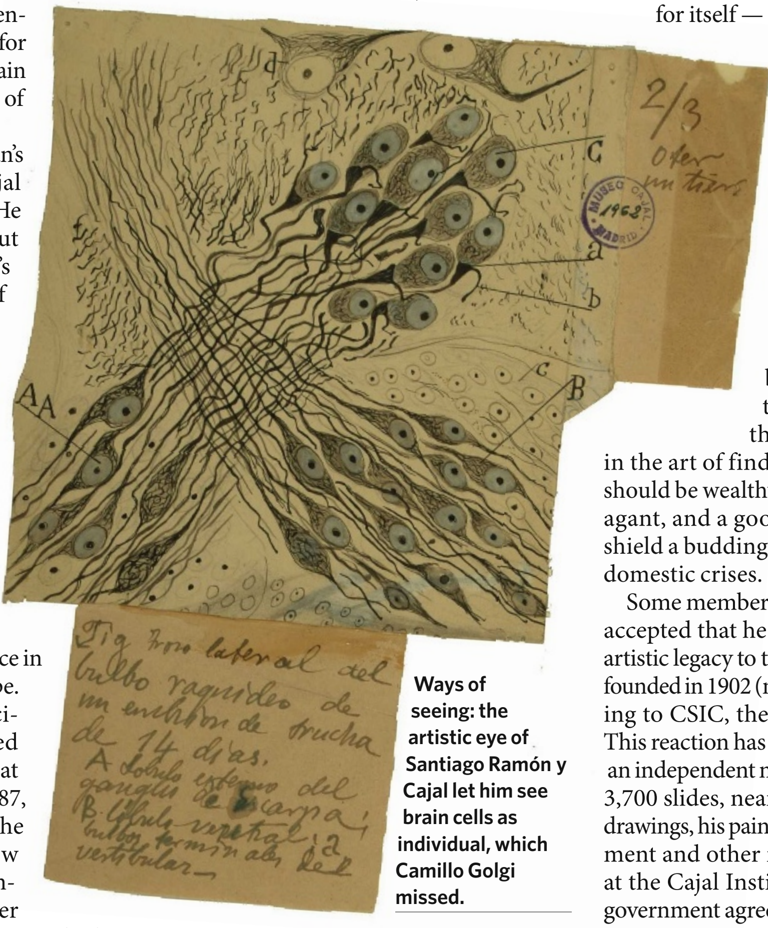
Spain was well off the beaten scientific path, but Cajal always tried to keep abreast of developments that might be useful to microscopy. In 1887, when he was already 35 years old, he travelled to Madrid to watch a new demonstration of a method of staining individual nerve cells with silver chromate. Known as the black reaction, it had been developed in 1873 by Golgi. One look down the microscope and Cajal was hooked. "All was sharp as a sketch with Chinese ink," he later wrote in his autobiography.

Cajal dedicated the rest of his life to neuroanatomy. Gone was the sociable and outgoing man: no more casino visits, no more social-club outings, no more time-wasting with his expanding family. He stopped playing with colour photography and even gave up painting, his first passion.

In Europe, the race was on to crack the brain's neuroanatomy, and Cajal wanted to win it. He improved the resolution of the Golgi method, and over the next 30 years he prepared and stained thousands of brains from various species in different stages of development. He acquired the best Zeiss microscope available and, drawing on his artistic talent, precisely transferred what he saw onto paper. Often he would make a composite drawing of several micrographs to present a clearer, very persuasive, scientific picture.

Following the stained cells from their bulbous bodies up through their long axons and dendrites, Cajal described the way each axon terminates near the cell body of the next. He concluded that the anatomy of the brain was indeed, as other scientists were starting to suspect, based on single cellular units, as had been shown for the rest of the body a few decades earlier. He further deduced that neurons receive information at one end and convey it in one direction to the junction with the next cell.

Cajal also made the first references to the brain's plasticity, a concept that has become widely appreciated only in the past 20 years. For example, he noted that the brains of people with mental disabilities contain fewer small spines on neuronal dendrites, the branches that deliver information to neuron bodies. He suggested that mental activity might influence spine number, and that the brain changes shape during development.




Ways of seeing: the artistic eye of Santiago Ramón y Cajal let him see brain cells as individual, which Camillo Golgi missed.

CAJAL COLLECTION, MADRID

In 1889, terrified that his scientific discoveries would be overlooked, Cajal packed up his prodigious collection of slides and drawings along with his precious microscope and embarked on a two-month tour of European scientific centres. The tour culminated in his attendance at the third meeting of the German Anatomical Society in Berlin. His work spoke for itself — which was just as well, given his weak grasp of foreign languages — and won over most sceptics to the individual neuron theory. Except Golgi. The intimidating Italian made sure he was out of town when Cajal passed through Pavia to visit him.

Cajal published more than 100 scientific papers. He also penned popular books and articles and 'how-to' manuals, including one that guided young scientists in the art of finding a wife. A potential wife should be wealthy, he advised, but not extravagant, and a good housekeeper who would shield a budding researcher from distracting domestic crises.

Some members of Cajal's family have never accepted that he left his entire scientific and artistic legacy to the institute in Madrid that he founded in 1902 (now the Cajal Institute, belonging to CSIC, the Spanish research council). This reaction has stymied all attempts to create an independent museum for his collection. His 3,700 slides, nearly 2,000 exquisite scientific drawings, his paintings and photographs, equipment and other memorabilia are still stored at the Cajal Institute. Last year, the Spanish government agreed to recognize the collection as being of national interest. Full cataloguing has just been completed, and a museum site should finally be chosen this year.

In the meantime, the collection can be viewed by appointment. While appreciating the clarity of the obviously independent neurons in Cajal's drawings, scientists should heed a warning from the story of Golgi's failure to see what Cajal saw though the microscope, using the very stain that he developed. Golgi was a scientific giant in his own right, but he blindly suspended reasonable judgement when it came to the work of his rival. ■

Alison Abbott is *Nature*'s Senior European Correspondent.

**The Cajal collection is at the Cajal Institute, 37 Avenida del Doctor Arce, 28002 Madrid, Spain (http://tinyurl.com/3dk76e).**

# NEWS & VIEWS

METABOLISM

# Food alert

Joshua P. Thaler and David E. Cummings

**The gut prevents nutrient overload during a meal by promoting satiety and enhancing insulin secretion. New findings show that nutrients in the gut also activate a neural circuit that increases insulin sensitivity.**

Eating is essential to life, yet its episodic nature necessitates physiological adaptations to avoid excesses or deficits in circulating fuels, especially glucose and lipids[1]. As the first point of contact with ingested food, the gastrointestinal tract is ideally positioned to initiate after-meal adaptations. Indeed, when nutrients are delivered into the gut, homeostatic mechanisms in place there are activated so that blood glucose levels are perturbed less than when nutrients are delivered directly into the blood. An established reason for this effect is that ingested nutrients stimulate the release of gut peptides called incretins, which enhance secretion of the hormone insulin, the main controller of blood glucose levels[2]. On page 1012 of this issue, Wang *et al.*[3] describe another gut-mediated mechanism that contributes to the regulation of glucose levels: a neural circuit, initiated in the intestine in response to nutrient sensing, that increases sensitivity to insulin.

Among the main products of dietary-fat digestion are long-chain fatty acids (LCFAs), which are cleaved from triglycerides by gastrointestinal enzymes. To investigate the effect of intestinal LCFAs on glucose homeostasis in rats, Wang and colleagues used the sophisticated method of pancreatic clamping, which allows the insulin sensitivity of specific tissues to be quantified. They found that infusion of calorically insignificant amounts of triglycerides directly into the animals' duodenum (the upper portion of the small intestine) markedly and rapidly increased insulin sensitivity.

Insulin not only promotes glucose uptake into tissues but also suppresses production by the liver of glucose derived from stored fuels; both actions lower blood sugar levels. Intestinal lipid infusions in Wang and colleagues' experiments specifically increased insulin sensitivity of the liver, reducing glucose output from this organ without affecting tissue glucose uptake. In unclamped rats, duodenal lipid infusions also contributed to glucose homeostasis, establishing the relevance of this mechanism to normal physiology.

How can lipids in the intestine trigger these systemic effects? The authors found that, initially, an LCFA metabolite called LCFA-CoA is sensed by the intestine. (The exact location

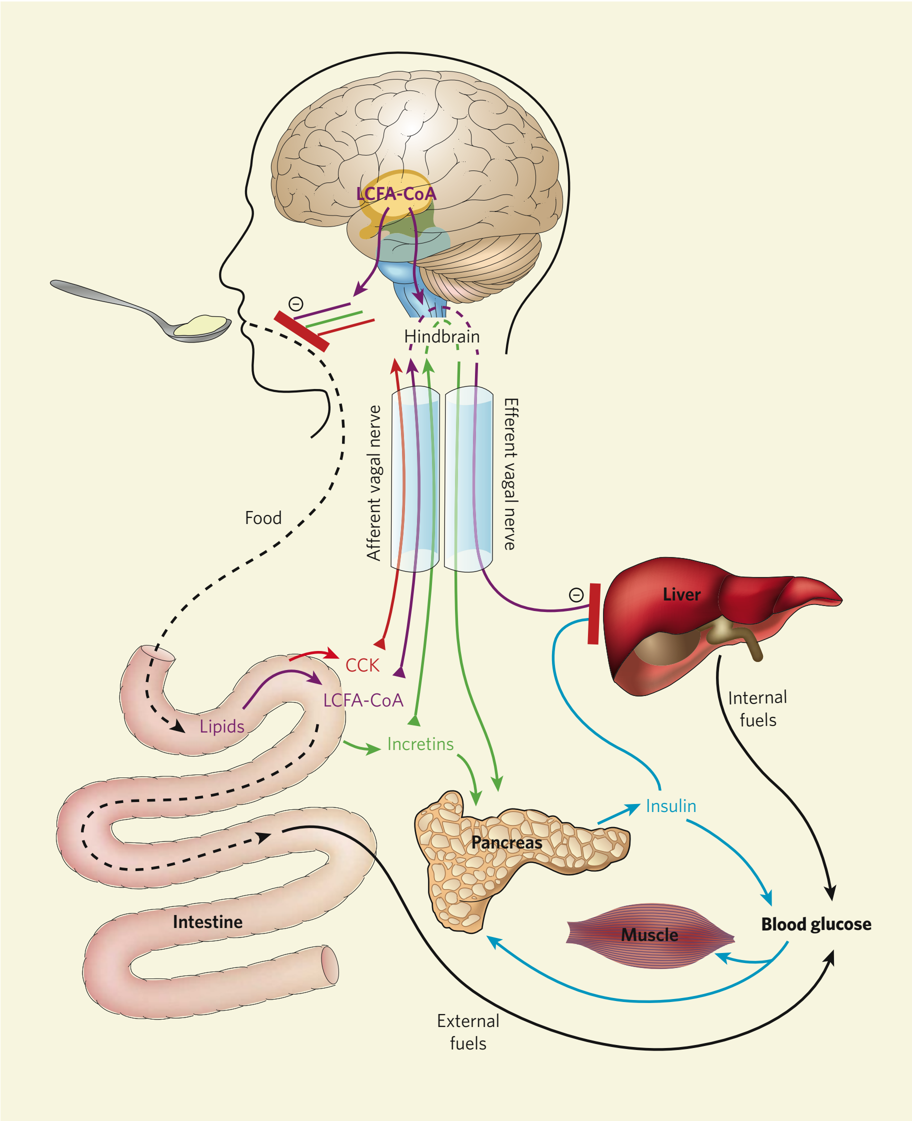



**Figure 1 | The intestine's contributions to glucose homeostasis.** Blood glucose comes from two sources: externally from ingested food, and internally from the mobilization of stored fuels by the liver. Following a meal, lipid sensing in the intestine initiates processes that limit both of these sources through several mechanisms. For example, the intestine secretes satiation peptides such as cholecystokinin (CCK) that promote meal termination, reducing further food intake (red pathways). Intestinal incretin peptides augment insulin secretion and also reduce food intake (green pathways). Wang *et al.*[3] now demonstrate that LCFA-CoA molecules in the intestine activate an intestine–brain–liver neural circuit that enhances insulin sensitivity, suppressing glucose output by the liver (purple pathways). In the brain, LCFA-CoA signalling also decreases both production of glucose by the liver and food intake. Thick red bars depict inhibition.

and identity of the intestinal sensor cells are unknown.) They also showed that the link between lipid sensing in the gut and insulin action in the liver involves an intestine–brain–liver circuit within the parasympathetic nervous system, a subdivision of the peripheral nervous system. The LCFA-CoA signal passes from the gut, along the vagus nerve to the brain, through the hindbrain, and then back down vagal nerve branches that terminate in the liver (Fig. 1). Wang and colleagues did not elucidate the details of communication between intestine and vagus, so intermediary roles for gut hormones, including incretins, remain possible.

They found that disruption of any component of this neural circuit eliminated the insulin-sensitizing effect of intestinal lipids, without affecting baseline glucose homeostasis. So the intestine–brain–liver axis serves not as a basal regulator of insulin sensitivity but as a first responder to meals, preventing the circulating nutrient excess that would occur with profligate mobilization of internal fuel stores following a meal.

Together with previous work, these findings[3] portray the gut as a rapid-response coordinator of energy input, influencing both the size of meals and the metabolic fate of the ingested nutrients. In response to food, the intestine produces satiety factors that promote meal termination[4]. The archetype, cholecystokinin, the production of which is stimulated by intestinal LCFAs, reduces food intake via an intestine–vagus–hindbrain pathway (Fig. 1). Furthermore, food consumption stimulates gut incretins to enhance insulin secretion. Lastly, Wang *et al.* show that LCFA sensing in the intestine also increases insulin sensitivity in the liver, minimizing inappropriate nutrient efflux from this organ after meals. So, more than just a calorie conduit, the intestine emerges as a neuroendocrine organ that regulates food intake as well as insulin secretion and action to improve glucose tolerance and to promote a seamless after-meal transition from fuel breakdown to storage.

Wang and colleagues' observations complement previous studies of LCFA signalling in the brain. As they do in the intestine, LCFA-CoA molecules in the hypothalamus activate neural pathways that increase insulin sensitivity in the liver, and they also reduce food intake[5–8] (Fig. 1). So LCFA-CoA seem to function acutely as a signal of nutrient abundance, triggering counter-regulatory responses that originate in the brain and the gut to limit further internal and external contributions to blood glucose levels. Unfortunately, both protective mechanisms are quickly inactivated by the ingestion of a fat-rich diet. For example, Wang *et al.*[3] found that the reduction in liver glucose production in response to intestinal lipids ceases after only three days of feeding rats a high-fat diet, and similar observations were made for the brain's LCFA-CoA sensor[5,6]. These results suggest that such diets promote obesity and diabetes in part by impairing nutrient-sensing systems that are designed to restrict food intake and enhance insulin sensitivity.

The apparent fragility of fatty-acid sensors in the intestine and the hypothalamus raises concerns about their 'real world' relevance. How important are fat sensors that can be overwhelmed by just three days of high-fat feeding? One hypothesis for the modern-day failure of mechanisms responsible for energy homeostasis posits that these systems evolved during periods of food scarcity and are thus ill-equipped for environments with abundant, palatable food. Moreover, the quantity and type of lipids in modern diets, typically high in saturated fats, might exert different effects from those observed by Wang *et al.*; the lipid infusions these authors used contained primarily polyunsaturated fats.

A chronic surfeit of dietary fat can also reverse the beneficial effects of the acute lipid infusions Wang and colleagues studied. Long-term exposure to dietary lipids increases fatty-acid oxidation, lowering LCFA-CoA levels. Moreover, chronic fat intake causes insulin resistance through weight gain and lipid accumulation in muscle cells. Lastly, compelling evidence indicates[9] that prolonged exposure to fatty acids from high-fat feeding and/or obesity stimulates inflammatory pathways that cause insulin resistance, perhaps overriding the acute insulin-sensitizing effects of intestinal lipids.

The revelation that intestinal nutrient sensing increases insulin sensitivity could aid our understanding of how bariatric surgery — operations that promote weight loss by modifying the gastrointestinal tract — ameliorates diabetes. A procedure known as Roux-en-Y gastric bypass surgery causes complete remission of diabetes in 84% of cases[10], and increasing evidence indicates[11] that this involves mechanisms beyond reductions in food intake and body weight. This operation, which alters the path of nutrients through the small intestine, increases the secretion of incretins. Wang and colleagues' work raises the untested possibility that complementary effects on the activity of the intestine–brain–liver neural circuit might further improve glucose metabolism.

The discovery of incretins led to a new class of antidiabetes drugs exemplified by exenatide. The LCFA-CoA-stimulated intestine–brain–liver circuit also provides potential targets for novel antidiabetes drugs and a conceptual basis for antidiabetes diets. ■

Joshua P. Thaler and David E. Cummings are in the Division of Metabolism, Endocrinology and Nutrition, Department of Medicine, Harborview Medical Center and VA Puget Sound Health Care System, 1660 South Columbian Way, S-111-Endo, University of Washington, Seattle, Washington 98108, USA.
e-mail: davidec@u.washington.edu

1. Woods, S. C. *Psychol. Rev.* **98,** 488–505 (1991).
2. Drucker, D. J. *Cell Metab.* **3,** 153–165 (2006).
3. Wang, P. Y. T. *et al. Nature* **452,** 1012–1016 (2008).
4. Cummings, D. E. & Overduin, J. *J. Clin. Invest.* **117,** 13–23 (2007).
5. Obici, S. *et al. Diabetes* **51,** 271–275 (2002).
6. Lam, T. K. *et al. Nature Med.* **11,** 320–327 (2005).
7. Obici, S., Feng, Z., Arduini, A., Conti, R. & Rossetti, L. *Nature Med.* **9,** 756–761 (2003).
8. Pocai, A., Obici, S., Schwartz, G. J. & Rossetti, L. *Cell Metab.* **1,** 53–61 (2005).
9. Wisse, B. E., Kim, F. & Schwartz, M. W. *Science* **318,** 928–929 (2007).
10. Buchwald, H. *et al. JAMA* **292,** 1724–1737 (2004).
11. Cummings, D. E., Overduin, J., Foster-Schubert, K. E. & Carlson, M. J. *Surg. Obes. Relat. Dis.* **3,** 109–115 (2007).

OPTICS

# Light reined in

Diederik Sybolt Wiersma

**Light always travels at the same speed in a vacuum, no more, no less. But in materials, there's room for manoeuvre: tweak the right material in the right way, and exciting optoelectronic properties result.**

In air, light travels at a speed of about 300,000 kilometres per second. That means it can circle Earth more than seven times in one second — so stupefyingly fast that we generally consider the arrival of light rays to be instantaneous. Finding ways to slow down and control the speed of light has inspired researchers for decades, both because it is fundamentally interesting and because it is potentially of practical use.

If you could moderate light's speed in a material well known for its superior optoelectronic properties — thus giving the light more time to interact and exchange energy with the material— you would be on to a winner. Writing in *Physical Review Letters*, Shubina *et al.*[1] describe how they have done just that in gallium nitride. This semiconductor has a wide bandgap such that it emits light at blue wavelengths; is easily 'doped' with impurities to create optoelectronic interactions; and is mechanically robust, even at high temperatures. Of the many uses gallium nitride has found in optoelectronic devices, the most cutting-edge is in the blue laser diodes used to read and write high-density Blu-ray storage discs[2]. This latest advance from Shubina and colleagues will aid the development of integrated components for optical signal processing, and potentially of detectors and solar cells that use gallium nitride.

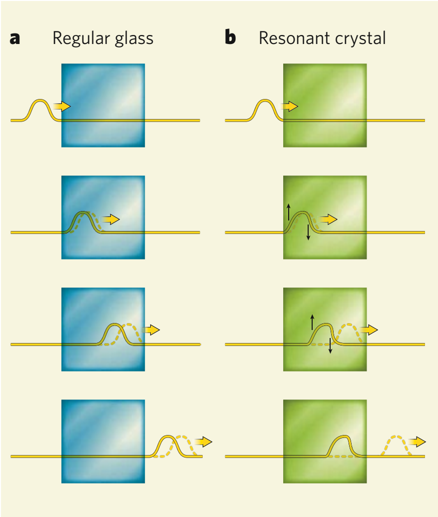



**Figure 1 | Reshaping at light speed. a,** When a pulse of light propagates through a block of a regular material with a constant refractive index — glass, for example — it is refracted and slightly delayed. **b,** In a strongly resonant material, by contrast, specific frequency components contained in the light pulse can be accelerated or delayed. As Shubina *et al.*[1] demonstrate in gallium nitride, if these changes act to suppress the leading edge (arrows) and increase the trailing part of the pulse, the maximum of the pulse envelope can be brought to a near standstill.

Controlling the speed of light is surrounded by many apparent mysteries and a fair sprinkling of misunderstandings. Admittedly, slowing light down until it stands still or even goes backwards does seem incredible. And, in the other direction, making light pulses travel faster than the speed of light in a vacuum is sometimes said to violate causality — wrongly, as it turns out.

Both effects are in fact intimately related. All one needs to do is bear in mind that a light pulse generally contains a superposition of light waves of different frequencies. The speed at which the whole envelope of the pulse is propagating is known as the group velocity[3]. Within this envelope, each individual frequency component of the pulse travels at its own phase velocity, which is determined by the refractive index of the material.

All one needs to do now is create a highly dispersive medium: that is, one whose refractive index depends strongly on the frequency of the passing light. On entering such a medium, different components of the pulse will begin to propagate at different phase velocities. The envelope of the pulse begins to change shape, and the top of the pulse might move forwards or backwards with respect to the rest (Fig. 1). In the first case, the pulse maximum seems to propagate at above vacuum light speed; in the second case, the effect can be so strong that the top of the pulse seems to stand still. But note that in the first case causality is not violated: the pulse itself, carrier of information, travels on as fast as ever.

A good way to achieve strong dispersion is by creating 'resonant' interactions that only affect light in a narrow frequency range. This can be achieved using a material with resonant microstructures, or one in which the constituent atoms or molecules have a resonant interaction with light of specific frequencies. This method has already been used to slow down light in crystals[4,5] and atomic gases[6], and Shubina *et al.*[1] now use it to put the brake on light in gallium nitride.

In crystals such as gallium nitride, light waves at specific frequencies can excite electrons, creating 'excitons' that temporarily store the optical energy[7]. This is strongly resonant behaviour — the interaction takes place only when the frequency of the light exactly matches that of the exciton — and thus strong dispersion is obtained[4]. When the authors shine a short light pulse with the right central frequency onto their sample, they observe a delay and elongation of the pulse, corresponding to a more than 100-fold reduction in velocity.

Shubina *et al.* discuss two possible mechanisms that might give rise to this effect, although they are not able to give a definitive picture of the physics behind their observations. First, the interaction of the light with the generated exciton might create an ' exciton–polariton' that delays the light's ballistic propagation through the material. Second, strong scattering effects around the exciton's resonant frequency could delay the light waves enormously[8]. Such effects have been observed in diffusive scattering from random distributions of uniform microspheres, and become enormous in cold atom clouds[9].

A further, much less intuitive delaying and trapping effect is that of optical localization[10]. Here, multiply-scattering light waves randomly create standing waves at points throughout the crystal. In strongly scattering semiconductors, the effect can become so strong that it brings diffusive propagation to a standstill[11]; it would be fascinating to see if it could be made to work in gallium nitride.

On a more immediately practical level, Shubina and colleagues' gallium nitride light 'brakers' could be applied in optical signal processing to produce delay lines for signal synchronization or to create short-lived optical memories[12]. This would require a method for tuning the resonance over a wider frequency window, not just at one specific exciton resonance frequency. Given gallium nitride's potential for integrating photonics and high-speed electronics, the effort would be worth while. One can imagine, for example, a photonic crystal created by periodically patterning a waveguide made of gallium nitride combined with other semiconductors, in which the gallium nitride would act both as a source of light and as delay components in a photonic circuit. ■

Diederik Sybolt Wiersma is at the European Laboratory for Nonlinear Spectroscopy (LENS) and INFM-BEC, Sesto-Fiorentino, 50019 Florence, Italy.
e-mail: wiersma@lens.unifi.it

# 50 & 100 YEARS AGO

## 50 YEARS AGO

Of the many craters on the Earth known to have been produced by fallen meteors, a few have left no signs of the meteor which caused them, apart from the huge holes created in the Earth's crust. Large meteorite fragments have been found in or near most craters. A possible explanation of the lack of meteorite fragments in the other cases is that the meteors concerned were of contraterrene constituency (anti-matter). In this case no traces of the meteors would remain, due to the annihilation process. One of these events, from which no fragments appear to have survived, has occurred in recent years. In 1908 a meteor, apparently of great size, was seen passing over western Russia; a few minutes later it crashed in the Siberian wastes with an explosion, effects of which were felt for hundreds of miles around, and seismographs thousands of miles away registered its impact ... No expedition to the site of its landing was undertaken until 1927, when the devastation found even at 40 miles from the site was almost inconceivable; but no fragments of the meteor were located despite the several huge though shallow craters which it had caused. If this meteor consisted of anti-matter an obvious explanation is at hand for the devastation found, the lack of meteoric material, and the great size reported in eye-witness accounts of its passage.
From *Nature* 26 April 1958.

## 100 YEARS AGO

At the London Institution on April 15, Mr Valdemar Poulsen lectured on "Telephoning without Wires." ... The progress made in wireless telephony is shown by the fact that conversation has been carried on across Denmark from Lyngby to another wireless telephone exchange at Esberg, 170 miles distant. The reproduction of the voice was clear and distinct, and easy to recognise.
From *Nature* 23 April 1908.

1. Shubina, T. V. *et al. Phys. Rev. Lett.* **100,** 087402 (2008).
2. Someya, T. *et al. Science* **285,** 1905–1906 (1999).
3. Garrett, C. G. B. & McCumber, D. E. *Phys. Rev. A* **1,** 305–313 (1970).
4. Chu, S. & Wong, S. *Phys. Rev. Lett.* **48,** 738–741 (1982).
5. Bigelow, M. S., Lepeshkin, N. N. & Boyd, R. W. *Science* **301,** 200–202 (2003).
6. Hau, L. V. *et al. Nature* **397,** 594–598 (1999).
7. Fleischhauer, M. & Lukin, M. D. *Phys. Rev. Lett.* **84,** 5094–5097 (2000).
8. van Albada, M. P., van Tiggelen, B. A., Lagendijk, A. & Tip, A. *Phys. Rev. Lett.* **66,** 3132–3135 (1991).
9. Labeyrie, G. *et al. Phys. Rev. Lett.* **91,** 223904 (2003).
10. John, S. *Phys. Rev. Lett.* **53,** 2169–2172 (1984).
11. Wiersma, D. S., Bartolini, P., Lagendijk, A. & Righini, R. *Nature* **390,** 671–673 (1997).
12. Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. *Nature* **409,** 490–493 (2001).


NEUROSCIENCE

# Current views on odour receptors

Alexander Chesler and Stuart Firestein

**Insects possess refined olfactory systems that use specific receptors on their antennae. It emerges that these receptors not only detect odour molecules but, unexpectedly, can also act as ion channels.**

All creatures sample their environment for chemicals that indicate the presence of food, mates, predators, dangers and attractions through mechanisms that were thought to be evolutionarily conserved from nematode worms to mammals. For example, in many organisms, odorant molecules bind to their receptors on the surface of neurons, initiating an intracellular signalling cascade. This ultimately results in the opening of ion channels conveying current, thereby completing the transformation of these chemical signals into electrical signals. It was thus assumed that the olfactory system of flies would function similarly. But when the fly odour receptor proteins were finally identified[1–3], a new riddle surfaced: what are the intermediary molecules that couple these receptors to the activation of ion channels in insect sensory neurons? Reporting in this issue, Sato *et al.*[4] and Wicher *et al.*[5] provide an answer to this question, but in doing so raise others that are just as puzzling.

To detect odours, vertebrates and worms use members of a large family of G-protein-coupled receptors (GPCRs), which thread through the cell membrane seven times. The initial identification of odorant receptors (ORs) in insects relied on the assumption that these organisms also detect odours through GPCRs. In 1999, three laboratories identified[1–3] a family of these 'seven-transmembrane' receptors whose expression in the antennae of the fruitfly *Drosophila melanogaster* turned out[6] to control the responses of individual olfactory neurons to chemical odorants.

But the first hint that insects detect odours differently from mammals came from analysis of the receptors' amino-acid sequences[1–3]. Although the vertebrate receptors are similar in sequence to other known GPCRs, the fly receptors are not. Moreover, unlike mammals, in which only one gene encoding a receptor is believed to be expressed in a given neuron, in fly sensory neurons at least one member of the receptor family Or83b is co-expressed with the insect odorant receptors. (Or83b is an essential co-receptor that does not bind to any known ligand, but is necessary for the intracellular transport and proper function of its companion receptor[7].) Finally, and perhaps most peculiarly, the insect receptors seem to lie in the membrane in the opposite configuration to GPCRs — that is, their amino terminus faces the cytoplasm and their carboxy terminus faces the outside[8].

The observations of Sato *et al.*[4] (page 1002) and Wicher *et al.*[5] (page 1007) may set things straight, albeit in an unexpected way. Both teams independently find that odour stimulation of multi-unit OR/Or83b complexes in cells grown in culture leads to inward cationic currents. Moreover, they show that the ionic conductance and permeability of the channels carrying these currents are similar to those of other known voltage-independent, non-selective channels for positive ions. Together, the authors' observations provide compelling evidence that, for odour detection, insects use an unusual strategy and an unusual receptor — the 'receptor' is in fact an ion channel. However, the two groups[4,5] have remarkably different mechanistic views of the sequence of events that occur from ligand binding to channel opening.

Through various experiments, Sato *et al.* rule out any role for G proteins. Instead, they conclude that the OR/Or83b complexes include an ion-channel function that is directly activated by odorant ligands. By contrast, Wicher and colleagues' data implicate a signalling molecule — the cyclic nucleotide cAMP — that is produced in response to G-protein-mediated signalling, which then activates the receptor/channel. These authors find that cAMP is generated in response to odours, and that modulators of G-protein signalling can affect OR/Or83b function. Most surprisingly, they also show that Or83b alone can form a cyclic-nucleotide-sensitive channel.

How can these contradictory findings be reconciled? One explanation might lie in the timescale of events that the two groups investigated. Sato *et al.* concerned themselves with the initial response that occurs within the first second or so after odour application, whereas Wicher *et al.* focused mainly on the cAMP-dependent responses that peak at between 30 and 60 seconds. So it could be that the opening of OR/Or83b complexes is regulated by both ligands and cyclic nucleotides, but on different timescales: odorants mediate fast detection by opening the channels, whereas cyclic nucleotides provide longer-lasting modulation. Although further work is needed to tease all this apart, the two studies add yet another unexpected twist to our knowledge of the seven-transmembrane receptor family: along with another GPCR channel, channelrhodopsin, and the structurally similar halorhodopsin ion pump[9,10], there is now an odour-activated channel receptor.

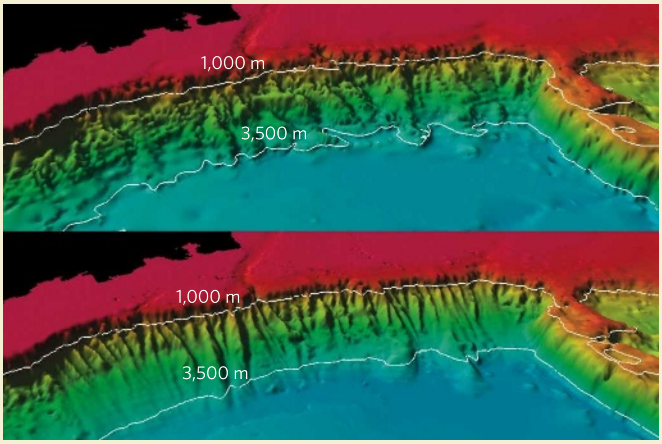
The coincidental arrival of these two papers and the subsequent review process raises an interesting set of editorial issues. We acted as a referee, and were somewhat involved in the process. Separately, each paper presents a convincing argument for a novel receptor-channel protein; taken together, they contain contradictory data that are difficult to reconcile. Of course, this is not the first time papers with contradictory views have been published. In the 1980s a controversy raged over the identity of the intermediate signalling molecule in photoreception — calcium or cGMP. No fewer than five papers appeared in *Nature* alone describing results that supported one or the other candidate, until a pivotal experiment[11] resolved the issue in favour of cGMP. Although journals, editors and reviewers have a responsibility to vet papers for accuracy and correctness, we must also recognize that, at its frontiers, science doesn't always produce unequivocal data. It is no less a responsibility to ensure that the literature accurately reflects what is known at the time. As George Bernard Shaw (paraphrasing Immanuel Kant) said in a toast to Albert Einstein: "Science is always wrong. It never solves a problem without creating ten more." Would we have it otherwise? ■

Alexander Chesler and Stuart Firestein are in the Department of Biological Sciences, Columbia University, 923 Fairchild Center, MC 2438, New York, New York 10027, USA. Alexander Chesler is at present in the Department of Physiology, University of California, San Francisco.
e-mail: sjf24@columbia.edu


1. Clyne, P. J. *et al. Neuron* **22,** 339–347 (1999).
2. Gao, Q. & Chess, A. *Genomics* **60,** 31–39 (1999).
3. Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. *Cell* **96,** 725–736 (1999).
4. Sato, K. *et al. Nature* **452,** 1002–1006 (2008).
5. Wicher, D. *et al. Nature* **452,** 1007–1011 (2008).
6. Hallem, E. A., Ho, M. G. & Carlson, J. R. *Cell* **117,** 965–979 (2004).
7. Larsson, M. C. *et al. Neuron* **43,** 703–714 (2004).
8. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. *PLoS Biol.* **4,** e20 (2006).
9. Nagel, G. *et al. Science* **296,** 2395–2398 (2002).
10. Lanyi, J. K. *Annu. Rev. Biophys. Chem.* **15,** 11–28 (1986).
11. Fesenko, E. E., Kolesnikov, S. S. & Lyubarsky, A. L. *Nature* **313,** 310–313 (1985).

OCEANOGRAPHY

# Bottom of the top of the world




The floor of the Arctic Ocean comes into sharper focus with the publication of an improved version of a bathymetric chart first released in provisional form in 1999, and as version 1 in 2001. Accurate mapping of the ocean bottom is essential for modelling deep ocean circulation, but also has a political angle in defining the extent of the continental shelf — a serious consideration in such a politically sensitive part of the world as the Arctic.

The story behind the improved bathymetric chart — IBCAO Version 2.0 — is told by Martin Jakobsson and colleagues in *Geophysical Research Letters* (M. Jakobsson *et al. Geophys. Res. Lett.* **35,** L07602; 2008). Its production is an instructive case of new data being married to a reinterpretation of old.

Most of the new data come from mapping missions carried out since 2000 with multibeam sonar equipment aboard various vessels, including USCGC *Healy*, RV *Polarstern* and IB *Oden*. Multibeam sonar systems differ from the sidescan systems used, for example, to look at the shape of the sea floor or to detect wrecks, in providing information mainly about depth.

The more dramatic changes to version 2 over version 1 are that, as the authors laconically put it, the "deep abyssal plains are systematically ca. 50–60 m deeper …". The revision stems from a metadata analysis of records collected by US Navy submarines over several decades, which are a central source of bathymetric information at high northern latitudes in particular. Conversion of data for version 1 was based on an assumption that the figure for the speed of sound in water used for the original calculations was 1,500 m $s^{-1}$. But in many cases the figure applied was 1,463 m $s^{-1}$. Hence the change in estimated depth, which also helps to explain several anomalies evident in version 1.

The three-dimensional views shown here are depictions of the Alaskan Slope and Northwind Ridge before (upper image) and after Jakobsson and colleagues' exercise in producing version 2. The image is about 650 km across, and the black area at the upper left is Alaska; the Northwind Ridge is the 'peninsula' on the right. The improved definition is evident in the sharper depiction of the gullies, caused by erosion, that scar the Alaskan Slope.

The new map is far from the final word. The authors point out that a near-perfect bathymetric model will require comprehensive multibeam coverage, which won't be available anytime soon. Meanwhile, more details on version 2 and derivations of it are available from www.ibcao.org.

**Tim Lincoln**

ASTROPHYSICS

# Exhaust inspection

David L. Meier

**What do you see if you peer into the exhaust of a jet engine larger than our Solar System? Only astronomers with the largest radio telescopes can see the full picture — and definitive observations are beginning to filter through.**

Cosmic jets — enormously energetic, highly directed beams of charged particles — can be ejected by stars as they form[1], as they die[2–4], and even, in certain instances, as they are reborn, rekindled by the accretion of surrounding gas[5,6]. The energy carried away by a jet can help to bring a dying massive star to explosion in a supernova, or disperse a less massive one as a planetary nebula. Jets provide some of the heating of the gas between stars, and can even disrupt the intergalactic medium in galaxy clusters.

Despite a profusion of hypotheses for how these cosmic power generators work, astronomers still don't have a clinching theory, largely because of a dearth of observations with which to compare models. New findings from Marscher *et al.* (page 966 of this issue)[7] go a long way to redressing this imbalance. The authors report how, in unprecedented detail, they imaged a jet engine belching out a puff of exhaust, allowing real-time insight into how jets work.

Theories exploring the role of gas accretion in jet acceleration originated in the late 1970s and early 1980s[8–10], and have more recently been extended to take account of features such as flows at close to the speed of light and the very hot temperatures of the jet plasma[11]. Many of the jet-forming objects are dense, accreting agglomerations of matter such as neutron stars or black holes, both of which originate from failed stars. But jets are also produced by objects such as fading 'white dwarf' stars, or even stars such as the Sun when they are first born. Understanding jet emission from one type of these objects, therefore, might lead to an understanding of many others.

The favoured picture that has emerged is of a mechanism similar to the operation of a fighter plane's jet engine, but confined within the lines of a strong magnetic field rather than a steel casing. Rotation of the magnetosphere around the central object creates a stiff, helical field that acts as both a turbine and a nozzle, expelling the exhaust flow and collimating it into a narrow beam (see Fig. 3 on page 968). Far from the object, this natural pinching action can overcollimate the beam, so that it starts to converge. The point of maximum convergence is known as the modified magnetohydrodynamic fast point; beyond that point, the jet can produce a strong 'fast-mode' shock wave that can distort or destroy the well-ordered structure of the magnetic nozzle. Other, less dramatic, 'slow-mode' shocks — pressure waves that travel parallel to the twisting field lines and do not disrupt them — can develop before the flow reaches the modified fast point.

The equations governing these processes are very complex, and several simplifying assumptions are needed to solve them. For that reason, the whole picture of magnetic jet acceleration just sketched is still a matter of debate. Marscher *et al.*[7] are the first to test the picture observationally. Their results, at least partially, confirm the model.

The authors' test-bed was the notoriously unruly BL Lacertae, the archetype of a particularly extreme form of jet emitter known as a blazar. BL Lacertae is an active galaxy containing a central supermassive black hole, and produces a jet with a speed 99% that of light. The jet, to within 10°, is pointing directly at us, so that looking down into the jet nozzle allows Earth-bound observers to examine the engine's inner workings directly.

This particular jet emits enormous amounts of energy at radio frequencies, and is huge — about two light years across. So, despite its immense distance from us (about 900 million light years), its angular size seen from Earth is fairly large, at around 0.5 milliarcseconds

(one arcsecond is 1/3,600th of a degree). Regular stellar jet engines in our Galaxy might be much closer, but they are diminutive by comparison, just nanoarcseconds in size. The jet from BL Lacertae was big enough for Marscher *et al.* to exploit a technique known as very-long-baseline interferometry (VLBI), which uses radio telescopes scattered around Earth's surface to create essentially one big radio telescope. In this case, these were the ten telescopes of the National Radio Astronomy Observatory's Very Long Baseline Array (VLBA), situated across the United States. Using this technique, the authors could make images of the jet with an angular resolution of about 0.1 milliarcseconds.

The VLBI technique, coupled with γ-ray, X-ray and optical-wavelength observations, allowed the authors to observe what happened as BL Lacertae sent a new pulse of plasma down its jet nozzle (see Fig. 1 on page 966). During the period of observation, the source twice sent out a flare of very-high-energy γ-rays. The first event looked very much like a slow-mode shock wave, tracing the plasma pulse as it travelled around one coil of the helical magnetic field. The field would thus seem to be well ordered up to that point, as predicted by theoretical models.

The second γ-ray flare coincided with a strong brightening of the radio 'core'; by this time, the pulse had probably left the magnetic nozzle and was passing through the strong fast-mode shock. Unfortunately, measurement limitations mean that we cannot be certain whether the magnetic field was helical or disordered in this core region during the second flare, although the strengthening and weakening of the radio polarization during and after the flare points to the second option. In that case, the fast-mode shock seems to have at least partially disrupted the helical field structure. Thus, despite the relative simplicity of the theoretical models, many of their predictions seem borne out by observations. This confirmation of the theoretical understanding of systems containing supermassive black holes increases the likelihood that jets of all sizes and types work in a similar way.

At present, jet acceleration from any cosmic object can be imaged only using the technique of VLBI at radio wavelengths. Observations at γ-ray wavelengths alone create the impression that flares may be produced within one-thousandth of a light year of a central black hole, but these latest observations[7] indicate that they occur more than a light year away. Thus, when NASA's γ-ray Large Area Space Telescope GLAST, which is due to launch on 16 May, starts to deliver jet measurements, it will need VLBI co-observations to deliver a complete picture. Japan's new VSOP-2 satellite will increase the radio imaging resolution of jets such as that of BL Lacertae by at least a factor of three, but to complete much of its proposed work it will need instruments such as the VLBA to supply detailed interferometry images[12]. Unfortunately, VLBI instruments are currently facing a funding gap that could see them scaled back or closed. The results of Marscher *et al.*[7] underscore how vital these instruments are to the understanding of jet acceleration throughout the Universe. ■

David L. Meier is at the NASA Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA.
e-mail: dlm@sgra.jpl.nasa.gov

1. Stapelfeldt, K. *et al. IAU Symp.* **182,** 355–364 (1997).
2. Sahai, R. *Astron. Soc. Pacif. Conf. Ser.* **313,** 141–147 (2004).
3. Wang, L. *et al. Astrophys. J.* **550,** 1030–1035 (2001).
4. Sari, R., Piran, T. & Halpern, J. P. *Astrophys. J.* **519,** L17–L20 (1999).
5. Mirabel, I. F. *C. R. Phys.* **8,** 7–15 (2007).
6. Bridle, A. H. & Perley, R. A. in *High Energy Astrophysics* (ed. Lamb, F.) 367–406 (Benjamin/Cummings, Menlo Park, 1985).
7. Marscher, A. P. *et al. Nature* **452,** 966–969 (2008).
8. Blandford, R. D. & Znajek, R. *Mon. Not. R. Astron. Soc.* **179,** 433–456 (1977).
9. Blandford, R. D. & Payne, D. G. *Mon. Not. R. Astron. Soc.* **199,** 883–903 (1982).
10. Meier, D. L., Koide, S. & Uchida, Y. *Science* **291,** 84–92 (2001).
11. Vlahakis, N. & Königl, A. *Astrophys. J.* **596,** 1080–1103 (2003).
12. Hagiwara, Y., Fomalont, E., Tsuboi, M. & Murata, Y. (eds) *Approaching Micro-Arcsecond Resolution with VSOP-2: Astrophysics and Technology* (Astron. Soc. Pacif., Seattle, WA, in the press).

GENE TRANSCRIPTION

# Two worlds merged

David M. Lonard and Bert W. O'Malley

**Why would two distant genes — on separate chromosomes and from different nuclear locations — unite in response to signals for gene expression? They might be seeds for the formation of transcriptional hubs.**

Gene transcription occurs largely at the submicroscopic scale. So although microscopic analysis of nuclear architecture has implicated various structures in this process[1], it has lacked the power to unravel the role that higher-order organization of chromatin (complexes of DNA and histone proteins) has in the expression of individual genes. Consequently, it has been difficult to combine whole-cell approaches such as microscopy with the molecular and biochemical techniques[2] that are primarily used to study gene expression. Fortunately, technological advances in both microscopy and molecular approaches are closing this gap. For instance, writing in *Cell*, Nunez *et al.*[3] bring a refreshing concordance between these two types of method to show that gene expression associated with activation of the nuclear receptor ERα (oestrogen receptor-α) depends in part on a large-scale reorganization of the genome that involves interactions both within and between chromosomes.

Much of the recent progress in understanding how gene transcription factors interact with chromatin can be credited to 'ChIP-on-chip' technology[4]. For example, this method revealed that ERα interacts with a specific set of DNA sequences known as oestrogen-response elements[5] (EREs), and that only some of these interactions enhance transcription. Other transcription-factor binding sites adjacent to EREs also seem to function cooperatively in facilitating transcription, indicating that various non-coding DNA sequences influence gene expression[5].

For their analysis of chromatin reorganization in response to the activation of ERα transcription, Nunez *et al.* used many sophisticated approaches, one of which — chromatin-conformation capture — is also related to ChIP-on-chip. This technique allows the detection of long-range interactions between genes, even those on different chromosomes[6]. Specifically, the authors aimed to characterize the interaction between two genes regulated by the oestrogen hormone: *TFF1* on chromosome 21 and *GREB1* on chromosome 2. They find that, in response to hormone treatment, the chromosomal regions (loci) containing these genes physically reach out for each other.

Nunez and colleagues also find that genes at other oestrogen-regulated loci interact with each other, indicating that interchromatin pairing of DNA sequences is a common feature of ERα-mediated transcription. What's more, previous work using another approach — assessing the effect that restricting genes' movement within the nucleus has on their transcription — has shown that other functionally related genes, such as those with roles in immunity or cellular differentiation, also make interchromatin contacts[3]. So it seems that interchromatin interactions are not restricted to a particular cellular process.

Interchromatin contact between *TFF1* and *GREB1* depends on the binding of ERα to an ERE. Moreover, these interactions are also influenced by the activity of other proteins, including chromatin remodelling factors and proteins that bind to the cytoskeletal protein actin; this indicates that passive movement is not sufficient for chromosomal reorganization[1]. Nunez *et al.*[3] add to the list of proteins known to mediate interchromatin interactions, showing that transcriptional coactivators such as SRC-3/AIB1, CBP, p300 and PBP are also involved in this process. Furthermore, they show that another co-regulatory protein — the

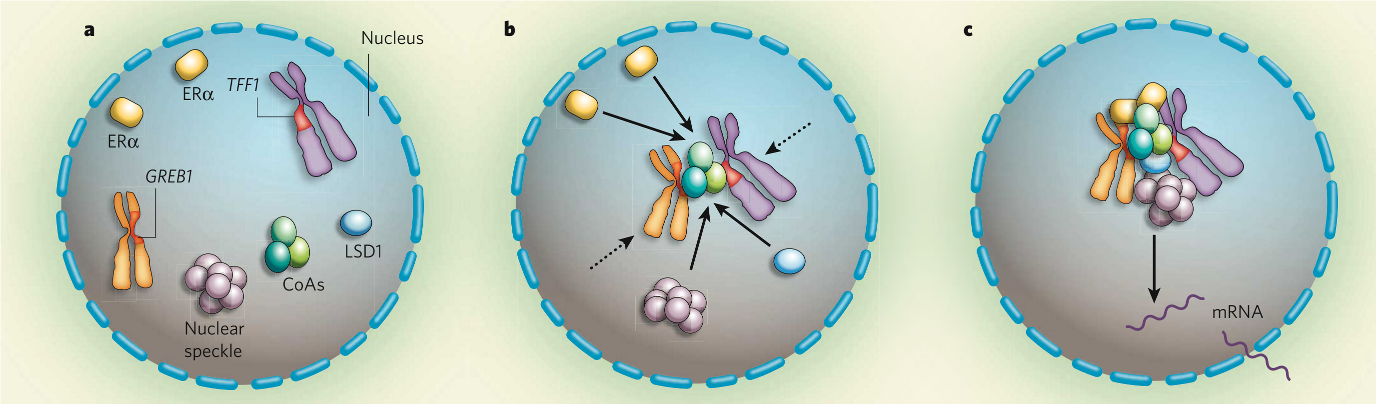



**Figure 1 | Rambling genes. a**, Nunez *et al.*[3] find that, in the absence of oestrogen, two genes (*TFF1* and *GREB1*) that are activated by this hormone and found on different chromosomes, reside in different locations within the nucleus. **b**, On exposure to oestrogen, these genes make interchromatin contacts, facilitated by the nuclear receptor ERα, transcriptional coactivators (CoAs), actin and molecular motor proteins. **c**, Subsequently, the gene pair interacts with a nuclear speckle in an LSD1-dependent manner, creating a multiprotein complex that might act as a transcriptional hub for DNA transcription into mRNA.

ERα-associated demethylase enzyme LSD1, which does not influence interchromosomal gene interaction — is required for the subsequent colocalization of the gene pairs with nuclear speckles (structures composed of high concentrations of specific proteins involved in messenger-RNA splicing), thereby probably leading to mRNA maturation (Fig. 1).

Transcriptional coactivators are multiprotein complexes that, collectively, have diverse enzymatic activities[7]. So it is tempting to speculate that, together with LSD1, distinct enzymes of a coactivator complex mediate actin polymerization (which might facilitate gene movement within the nucleus), interchromatin interactions and chromatin interface with transcription 'factories' — structures that generate mRNAs at higher rates. Ultimately, it is likely that coactivators and other receptor-associated proteins work together to orchestrate the formation of an ERα-activated transcriptional supercomplex comprising ERα itself, coactivators, mRNA-splicing factors, transcription factories and two or more oestrogen-regulated genes (Fig. 1).

The possibility of ERα-regulated genes converging into distinct hubs gives rise to some interesting questions. How common is the above process among other genes regulated by nuclear receptors? What signals define how one gene finds its appropriate partner gene? Do these transcriptional hubs contain functionally related genes, or is their organization simply spatial? Could it be that sequences close to a gene, or possibly specific chromatin modifications at each ERα-regulated locus, provide proper pairing information?

Many EREs are located in stretches of DNA between genes, far away from known coding sequences. Do these sites also engage with interchromatin loci pairs, and, if so, to what extent do they influence the transcription of the partner locus? Do large-scale structural rearrangements induced by oestrogen, or other signals that activate oestrogen receptors, affect genome-wide expression of genes that are not directly regulated by ERα? These are some of the intriguing questions that remain. But with the availability of sophisticated techniques such as those used by Nunez *et al.*[3], there is hope that the mysterious relationship between higher-order nuclear architecture, chromatin organization and gene expression will soon be revealed. ■

David M. Lonard and Bert W. O'Malley are in the Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.
e-mail: berto@bcm.tmc.edu

1. Kumaran, R. I., Thakar, R. & Spector, D. L. *Cell* **132,** 929–934 (2008).
2. Speicher, M. R. & Carter, N. P. *Nature Rev. Genet.* **6,** 782–792 (2005).
3. Nunez, E. *et al. Cell* **132,** 996–1010 (2008).
4. Nègre, N., Lavrov, S., Hennetin, J., Bellis, M. & Cavalli, G. *Methods Enzymol.* **410,** 316–341 (2006).
5. Carroll, J. S. & Brown, M. *Mol. Endocrinol.* **20,** 1707–1714 (2006).
6. Simonis, M., Kooren, J. & de Laat, W. *Nature Methods* **4,** 895–901 (2007).
7. Lonard, D. M. & O'Malley, B. W. *Mol. Cell* **27,** 691–700 (2007).

PLANT BIOLOGY

## Scent of a rose

CUBOIMAGES SRL/ALAMY

What Chinese roses lack in resistance to cold and disease, they make up for with repeated blooming and a distinctive fragrance. It isn't surprising, then, that plant breeders, fascinated with this ancient rose family, brought it to Europe in the nineteenth century to generate the now popular hybrid tea roses, among many others. But how did the distinctive scent of Chinese roses evolve? Gabriel Scalliet and colleagues now provide the answer (G. Scalliet *et al. Proc. Natl Acad. Sci. USA* **105,** 5927–5932; 2008).

Unlike their European counterparts, the main scent component of Chinese roses and their descendants is dimethoxytoluene. In the final steps of its formation, two methyl groups are sequentially added to a precursor molecule in reactions catalysed by enzymes called OOMT1 and OOMT2. These enzymes are almost identical in amino-acid sequence, yet they target different substrates.

Scalliet and colleagues pinpoint the crucial amino-acid residues of the enzymes that confer substrate specificity: a tyrosine in OOMT1 and a phenylalanine in OOMT2. Swapping these single residues between the two enzymes switched their substrate specificity, most probably by changing the steric and hydrophobic properties of their substrate-binding sites, where these residues reside. Moreover, OOMT2 could methylate the target of OOMT1 *in vitro*, albeit with lower efficiency.

Two nearly identical enzymes with nearly identical activities, that can both perform similar reactions, although together they are more efficient than alone — these are hallmarks of the products of duplicated genes, which arise when a gene doubles in number and one of the copies then undergoes a minor mutation. Indeed, Scalliet *et al.* show that, although all of the 13 European rose species they examined carry an OOMT-like gene, only the Chinese species, which are evolutionarily younger, carry two types of this gene. The authors' further phylogenetic analysis strongly hints that, in Chinese species, the second OOMT gene arose through a duplication event.

Roses are not unique in acquiring new functions through gene duplication; this is thought to be a fundamental mechanism for generating diversity between and within organisms. But it isn't always possible to trace such changes, which makes Scalliet and colleagues' work of interest not only to plant biologists but also to evolutionary scientists.

**Sadaf Shadan**

ARTICLES

# The genome of the model beetle and pest *Tribolium castaneum*

*Tribolium* Genome Sequencing Consortium*

**Tribolium castaneum is a member of the most species-rich eukaryotic order, a powerful model organism for the study of generalized insect development, and an important pest of stored agricultural products. We describe its genome sequence here. This omnivorous beetle has evolved the ability to interact with a diverse chemical environment, as shown by large expansions in odorant and gustatory receptors, as well as P450 and other detoxification enzymes. Development in *Tribolium* is more representative of other insects than is *Drosophila*, a fact reflected in gene content and function. For example, *Tribolium* has retained more ancestral genes involved in cell–cell communication than *Drosophila*, some being expressed in the growth zone crucial for axial elongation in short-germ development. Systemic RNA interference in *T. castaneum* functions differently from that in *Caenorhabditis elegans*, but nevertheless offers similar power for the elucidation of gene function and identification of targets for selective insect control.**

By far the most evolutionarily successful metazoans[1], beetles (Coleoptera) can luminesce (fireflies), spit defensive liquids (bombardier beetles), visually and behaviourally mimic bees and wasps, or chemically mimic ants that detect intruders by their foreign odour. Many beetles (for example, boll weevil, corn rootworm, Colorado potato beetle and Asian longhorn beetle) are associated with billions of dollars of agricultural and natural resource losses.

The red flour beetle, *Tribolium castaneum*, found wherever grains or other dried foods are stored, has a highly evolved kidney-like cryptonephridial organ to survive such extremely dry environments. It has demonstrated resistance to all classes of insecticides used against it. Like all beetles, *Tribolium* has elytra (wing covers) that coordinate precisely with folding wings, allowing flight while providing protection.

*Tribolium* facilitates genetic analysis with ease of culture, a short life cycle, high fecundity, and facility for genetic crosses (see ref. 2), allowing efficient genetic screens by means of chemical mutagens, radiation and binary transposon systems[3]. As in *Caenorhabditis elegans*, RNA interference (RNAi) is systemic in *Tribolium*, facilitating knockdown of specific gene products in any tissue, developmental stage or offspring of double-stranded (ds)RNA-injected females[4,5].

Particularly favoured for developmental studies, *Tribolium* is much more representative of other insects than is *Drosophila*[6]. In contrast to *Drosophila*, *Tribolium* larvae display eyes in a fully formed head and three pairs of thoracic legs (Supplementary Fig. 1). In addition, *Tribolium* develops via short-germ embryogenesis where additional segments are sequentially added from a posterior growth zone (Supplementary Fig. 1). This proliferative mechanism of segmentation differs from the *Drosophila* model, but resembles that of vertebrates and basal arthropods such as millipedes[7].

## Genome sequence and organization

Approximately 1.52 million sequence reads (7.3× coverage) were generated from the highly inbred Georgia 2 (GA2) strain and assembled into contigs totalling 152 megabases (Mb) and scaffolds spanning ~160 Mb of genomic sequence (Supplementary Tables 1–4 and Supplementary Information). Almost 90% of this sequence was mapped to the ten *Tribolium* linkage groups using a genetic map of ~500 markers generated from the GA2 strain[8]. Excluding heterochromatic regions dense in highly repetitive sequences, the genome is well represented and of high quality (see Supplementary Data for details).

**G+C content.** *Tribolium*, like *Apis*, has a very (A+T)-rich genome (33% and 34% G+C, respectively), but *Tribolium* G+C domains lack the extremes of G+C content present in *Apis mellifera* (Fig. 1 and Supplementary Fig. 3). Despite global G+C similarity to *Apis*, genes in *Tribolium*, as in *Anopheles* and *Drosophila* but not *Apis*, show a bias towards occurring in (G+C)-rich regions of the genome (Fig. 1). Whatever mechanism drives the accumulation of A+T nucleotides in *Tribolium*, it does not affect genes in the manner observed in the honeybee, where perhaps additional mechanisms are present.

**Repetitive DNA.** Fully one-third of the *Tribolium* genome assembly consists of repetitive DNA, which is also (A+T)-rich. Compared to other insects, there is a paucity of microsatellites (1–6-base-pair (bp) motifs) in *Tribolium*[9]. However, *Tribolium* contains a relative excess of larger satellites, including several with repeat units longer than 100 bp (2.5% of the *Tribolium* genome compared with 0.7% in *Drosophila*). Most (83%) of the microsatellites are found in intergenic regions (63%) or introns (20%), but there is strong over-representation of non-frameshift-causing repeats (3- and 6-bp motifs) due to a dearth of dinucleotide repeats (see Supplementary Information). Of 981 randomly chosen microsatellites, 509 (55.2%) are polymorphic in a sample of 11 *Tribolium* populations from around the world[9], providing an extensive collection of markers for population studies. Preliminary efforts to assess global population structure show a shallow but significant correlation between geographic and genetic distance (Supplementary Fig. 4). This suggests that anthropogenic dispersal may maintain a modest level of gene flow across vast distances in this human commensal.

**Transposable elements.** Transposable elements and other repetitive DNA accumulate in regions along each linkage group that resemble the pericentric blocks of heterochromatin visible in *Hpa*II-banded chromosomes[10]. These regions are probably composed largely of highly repetitive heterochromatic sequences, and represent most of the 44-Mb difference between the estimated genome size (0.2 pg or

*Lists of participants and affiliations appear at the end of the paper.

204 Mb[11]) and the current assembly (160 Mb). Indeed, as much as 17% of the *Tribolium* genome is composed of a 360-bp satellite[12] that constitutes only 0.3% of the assembled genome sequence. Several families of DNA transposons, as well as long terminal repeat (LTR) and non-LTR retrotransposons, constituting approximately 6% of the genome, were identified via encoded protein sequence similarity to previously identified elements using TEPipe or BLAST, and are listed in Supplementary Table 5.

**Telomeres.** *Tribolium* has a telomerase and telomeres containing TCAGG repeats[13], a variant of the standard arthropod TTAGG telomeric repeat. Manual assembly of the proximal regions of multiple telomeres beyond the ends of the assembled scaffolds (Supplementary Information) reveals TCAGG repeats interrupted by full-length and 5′-truncated non-LTR retrotransposons belonging to the R1 clade, best known for insertions in the rDNA locus[14]. *Tribolium* telomeres range in length from 15 kilobases (kb) upwards and probably represent a stage intermediate to the loss of telomeres and telomerase in Diptera compared with the simple canonical structure of the honeybee[15] or the more regular insertion of non-LTR retrotransposons into the simple repeats of the silkmoth[16].

## Gene content and the proteome

**Comparative gene content analysis.** To understand the consensus set of 16,404 gene models in the context of other available insect and vertebrate genomes, all genes were classified according to their degree of similarity using systematic cross-species analysis. Five insects (*Drosophila melanogaster*, *Anopheles gambiae*, *Aedes aegypti*, *T. castaneum*, *A. mellifera*) and five vertebrates (*Homo sapiens*, *Mus musculus*, *Monodelphis domestica*, *Gallus gallus*, *Tetraodon nigroviridis*) with similar phylogenetic branching orders were chosen for the comparison. We found the fractions of universal and insect-specific orthologues in *Tribolium* similar to other insect genomes, as expected, whereas the number of genes without similarity is considerably higher (Fig. 2), possibly attributable to less stringent gene prediction.

Over 47% of *Tribolium* genes (7,579) are ancient, with traceable orthologous relations between insects and vertebrates including 15% (2,403) universal single-copy orthologues. Another 1,462 *Tribolium* genes (9%) constitute the core of what are currently insect-specific orthologues. In comparison, 21% (4,937) of human genes have vertebrate-specific orthologues.

Several hundred ancient genes seem to be under limited evolutionary selection and were independently lost in several species studied (the patchy fraction, defined in Fig. 2). Each new genome uncovers previously invisible ancestral relations among genes—for example, as many as 126 orthologous gene groups shared between *Tribolium* and humans seem to be absent from the other sequenced insect genomes (Fig. 3 and Supplementary Table 10), 44 of which are single-copy genes present in all vertebrates.

The evolutionary emergence of many predicted *Tribolium* genes is not clear. Thousands of genes currently appear to be species-specific as either no sequence similarity to other genes is detectable, or homology but not orthology can be determined. Reassuringly, this fraction is similar in *Tribolium* and *Drosophila*.

We quantified the species phylogeny using a maximum likelihood approach with the concatenated multiple alignment of 1,150 universal single-copy orthologues present in all the organisms studied—an ideal genome-wide data set of essential genes evolving under similar constraints (Fig. 2 and Supplementary Fig. 6). This analysis confirmed previous analyses based on expressed sequence tag (EST) sequences that the Hymenoptera are basal within the Holometabola[17]. The shorter branch length for *Tribolium* implies that the elevated rate of evolution observed in *Drosophila* and *Anopheles* occurred more recently[18].

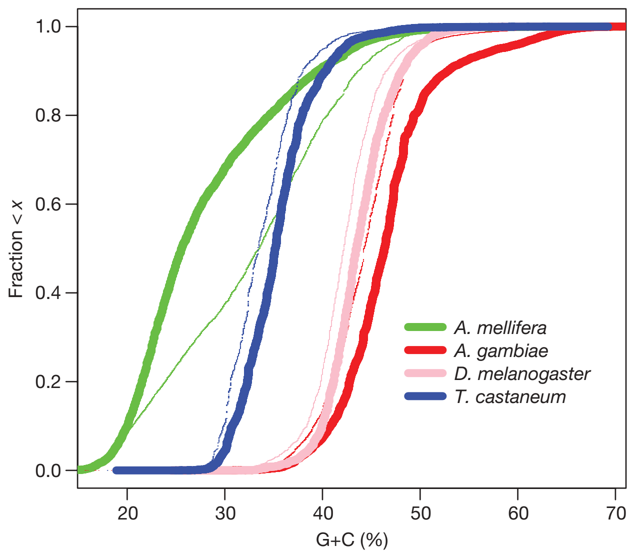



**Figure 1 | Cumulative distribution of genic and genomic G+C-content domains in *Apis mellifera*, *Anopheles gambiae*, *Drosophila melanogaster* and *Tribolium castaneum*.** Cumulative distributions show the fraction of genes (thick lines) or of the entire genome (thin lines) occurring in G+C-content domains less than a given percentage G+C (<*X*). The more (A+T)-rich half of the *T. castaneum* genome contains only 30.8% of all *T. castaneum* genes (31.4% and 33% of *A. gambiae* and *D. melanogaster* genes, respectively), whereas the more (A+T)-rich half of the *A. mellifera* genome contains 77.6% of its genes. At every point on the *T. castaneum*, *A. gambiae* and *D. melanogaster* curves there are fewer genes present in the fraction of the genome less than a given percentage G+C than would be expected if the genes were randomly distributed. In contrast, *A. mellifera* exhibits the opposite distribution.

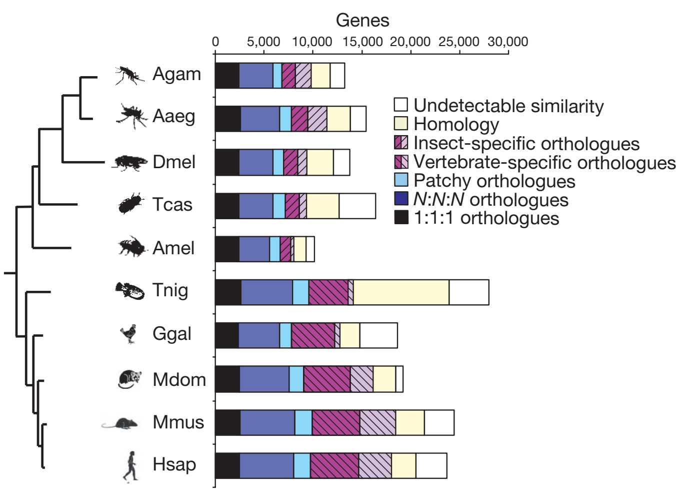



**Figure 2 | Insect gene orthology.** Comparison of the gene repertoire in five insect and five vertebrate genomes, ranging from the core of metazoan genes (dark blue fraction on the left) to the species-unique sequences (white band on the right). The striped boxes correspond to insect- and vertebrate-specific orthologous genes, where the darker bands correspond to all insects or vertebrates (allowing one loss). *N*:*N*:*N* indicates orthologues present in multiple copies in all species (allowing one loss); patchy indicates ancient orthologues (requiring at least one insect and one vertebrate gene) that have become differentially extinct in some lineages. The species tree on the left (shown in detail in Supplementary Fig. 6) was computed using the maximum-likelihood approach on concatenated sequences of 1,150 universal single-copy orthologues. It shows an accelerated rate of evolution in insects and confirms the basal position of the Hymenoptera within the Holometabola[17]. Aaeg, *Aedes aegypti*; Agam, *Anopheles gambiae*; Amel, *Apis mellifera*; Dmel, *Drosophila melanogaster*; Ggal, *Gallus gallus*; Hsap, *Homo sapiens*; Mdom, *Monodelphis domestica*; Mmus, *Mus musculus*; Tcas, *Tribolium castaneum*; Tnig, *Tetraodon nigroviridis*.

Gene family expansion, frequently associated with a particular adaptation pressure, might reveal physiologically and phenotypically unique features of beetles (Supplementary Table 9 and protein family discussions below). Many duplications shaped the gene content of *Tribolium*, most notably among odorant-binding proteins and the CYP450 subfamilies CYP6 and CYP9 (Supplementary Fig. 11), some of which are involved in the development of insecticide resistance in the Diptera[19]. Duplication of genes under copy-number selection in other species is indicative of species-specific neo-functionalization[20]. At least 152 genes duplicated in *Tribolium* have single-copy status in all other insects studied, including sevenfold duplication of genes orthologous to *Drosophila* CG1625, encoding a putative structural constituent of cytoskeleton, and human ENSP00000269392, encoding centrosomal pre-acrosome localization protein 1.

We also analysed the phylogenetic distribution of orthologous gene group members to quantify evolutionary gene losses[21]. Although least affected, dozens of single-copy orthologues seem to be lost in each lineage. Thirty-eight such genes lost in *Tribolium* include rather unique genes, encoding phosphotriesterase-related protein and peroxisome assembly factor 1 (peroxin-2), compared to 59 such genes lost in *Drosophila*. Notably, for the less restricted fractions of orthologues (defined in Fig. 2), several hundred gene orthologues have been lost in each species.

## Analysis of specific gene sets

In addition to a global automated analysis of the predicted *Tribolium* gene set, the consortium manually annotated and analysed ~2,000 genes (some additionally subjected to RNAi and expression analysis), focusing on developmental processes and genes of importance for agriculture and pest management.

## Development

We identified and analysed homologues of known insect and vertebrate developmental genes to gain novel insights into the molecular basis of developmental differences between *Drosophila* and *Tribolium*. Supplementary Table 11 lists selected *Tribolium* developmental genes and their *Drosophila* and *Apis* orthologues.

**Oogenesis.** Despite profound differences in ovarian architecture—telotrophic versus polytrophic—we identified *Tribolium* orthologues

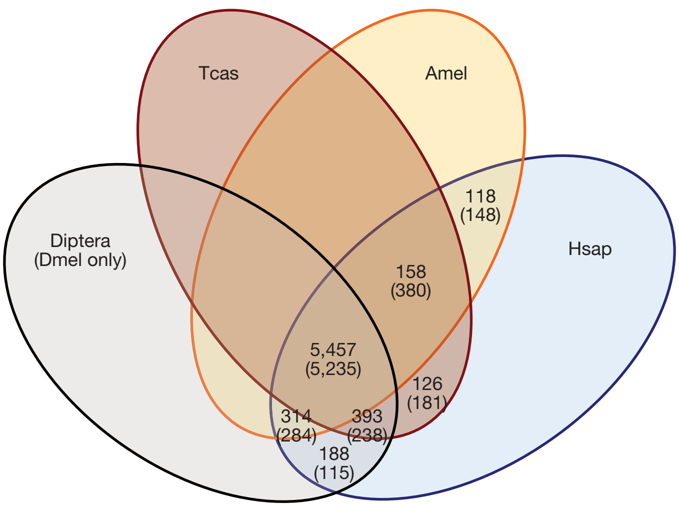



**Figure 3 | Orthologous genes shared between insect and human genomes.** The Venn diagram shows the number of orthologous groups of genes shared between the insect and human genomes. In addition to the majority of Urbilateria (last common ancestor of the Bilateria) genes shared by all the organisms, there are hundreds of genes that have been lost in some lineages (for example, only retained between human and *Tribolium* or human and honeybee, but lost in Diptera). Diptera is represented here by *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster* (with numbers considering only *D. melanogaster* shown in parentheses).

of most *Drosophila* genes required for stem cell maintenance, RNA localization and axis formation. Like *Apis*, however, *Tribolium* lacks a *bag of marbles* orthologue, which is essential for the differentiation of cystoblast versus germline stem cells in *Drosophila*[22]. Interestingly, an orthologue of the gene *gld-1*, which fulfils a similar function in *C. elegans*[22], is present in *Tribolium*.

**Anterior–posterior patterning.** Analysis of the genome sequence confirmed the absence of a *bcd* orthologue in *Tribolium*. Instead, anterior patterning is synergistically organized by *otd* and *hb* (ref. 23). However, it is still unclear how the posterior gradient of *Tribolium* Caudal is shaped in the absence of Bicoid. Notably, *Tribolium* contains an orthologue of *mex-3*, a factor that translationally represses the *C. elegans cad* homologue[24]. Although the *Tribolium* genome contains orthologues of the *Drosophila* segmentation genes, their functions are not entirely conserved[25–27]. Furthermore, the genome reveals the unexpected polycistronic organization of a novel gap gene, *mille-pattes*[28], the transcript of which encodes several short peptides.

In contrast to the classical protostomian model organisms *Drosophila* and *Caenorhabditis*, *Hox* genes in *Tribolium* map to a single cluster of ~750 kb on linkage group 2. Orthologues of all *Drosophila Hox* genes and the *Hox*-derived genes *ftz* and *zen* are transcribed from the same strand, and we find no evidence for interspersion of other protein-coding genes. Taken together, these results suggest that the evolutionary constraints preserving *Hox* cluster integrity still function in *Tribolium*.

**Dorso-ventral patterning.** As in *Drosophila*, the dorso-ventral axis of the *Tribolium* embryo depends on a nuclear gradient of Dl, an NF-κB protein, which is established through ventral activation of a Tl receptor[29] (one of four in *Tribolium*). Factors required for localized Tl activation are also present in *Tribolium* (potential Tl ligands: six *spz*-like genes; extracellular proteases: one *gd*, six *snk* and four tandem *ea* genes), suggesting that, as in *Drosophila*, an extra-embryonic signal induces the embryonic dorso-ventral axis.

*Tribolium sog* inhibition of Dpp/BMP generates a patterning gradient along the dorso-ventral axis[30]. Similar *chordin/sog* function in spiders and a hemichordate suggest that this may represent the ancestral bilaterian condition[30]. Like *Apis*, *Tribolium* lacks an orthologue of *Drosophila scw*, but knockdown of another ligand, *Tribolium gbb1*, affected the embryonic Dpp/BMP gradient. *Tribolium* contains orthologues of all five *Drosophila* TGF-β receptors; however, Dpp signalling moderators that have duplicated and diverged in *Drosophila*, such as *Tol/tok* and *Cv/tsg*, occur as a single copy in *Tribolium*. Most strikingly, *Tribolium* contains homologues of *BMP10* as well as *bambi*, *Dan* and *gremlin* BMP inhibitors, which are all known from vertebrates, but are not found in *Drosophila*.

**The growth zone.** We identified several members of the Fgf and Wnt signalling pathways. The expression patterns of *Tribolium Fgf8*, Wnt1, Wnt5 and WntD/8 (refs 31, 32) highlight the dynamic organization of the growth zone and underline its role in axis elongation.

**Head patterning.** Orthologues of 25 out of 30 key regulators of the vertebrate anterior neural plate are specifically expressed in the *Tribolium* embryonic head (Supplementary Table 12). Two orthologues are not expressed in the head neuroectoderm (*barH*, *arx*) and three do not have *Tribolium* or *Drosophila* orthologues (*vax*, *hesx1*, *atx*). Of the canonical *Drosophila* head gap genes, only the late head-patterning function of *otd* is conserved. *ems* function is restricted to parts of the antennal and ocular segments, and knockdown of *btd* seems to have no phenotypic consequences. Thus, analysis of *Tribolium* genes defines a set of genes that is highly conserved in bilaterian head development, and underscores the derived mode of *Drosophila* head patterning.

**Leg and wing development.** In contrast to *Drosophila*, ventral appendages in *Tribolium* develop during embryogenesis from buds that grow continuously along the proximo-distal axis[33]. Nonetheless, we identified *Tribolium* orthologues for most of a core set of *Drosophila* appendage genes (Supplementary Table 13). On the other

hand, orthologues of genes not found in *Drosophila*, such as *Wnt11*, *gremlin*, *Fgf8* and an F-Box gene, are expressed in the embryonic legs[21,31]. Although their exact function in *Tribolium* appendages is not known, Fgf8 is essential to vertebrate limb development.

A major innovation driving the radiation of beetles was the evolution of a highly modified protective forewing. Expression analysis and RNAi experiments revealed a high degree of conservation between *Tribolium* and *Drosophila* wing gene networks (Supplementary Table 13), supporting the hypothesis that sclerotized elytra evolved from ancestral membranous wings mainly through new interactions between conserved patterning modules and as yet unknown downstream effector genes.

**Eye development.** *Tribolium* has orthologues of nearly all genes currently known to regulate specification and differentiation in the *Drosophila* retina (Supplementary Table 14). Exceptions are the linker protein Phyllopod and the lens crystallin protein Drosocrystallin, which are restricted to Diptera. Eight of fifty-seven investigated eye developmental genes are duplicated in the *Drosophila* genome but not *Tribolium*, and in four cases the *Drosophila* paralogues have similar function. This suggests a more dynamic evolution of *Drosophila* retina genes and higher genetic complexity, highlighting the value of *Tribolium* as a more ancestral and simply organized model of insect eye development.

## Genes relevant to pest and *Tribolium* biology

*Tribolium castaneum* is a notorious invader of stored grains and grain products. Resultantly, much effort and expense is directed to find better ways to control this and other grain pests. Here we describe established and possible future pesticide targets, as well as genes underlying vision and taste. Finally, we describe genes forming the basis of systemic RNAi in *Tribolium*.

### Established insecticide targets

**Cys-loop ligand-gated ion channels.** Members of this superfamily mediate chemical synaptic transmission in insects and are targets of successful pest control chemicals with animal health and crop protection applications[34]. The *Tribolium* Cys-loop ligand-gated ion channel (Cys-loop LGIC) superfamily contains 24 genes, the largest known so far for insects (*Drosophila* and *Apis* superfamilies comprise 23 and 21 genes, respectively), due in part to the additional nicotinic acetylcholine receptor (nAChR) subunits in *Tribolium*. We also found genes for ion channels gated by γ-aminobutyric acid (γ-aminobutyric acid receptors (GABARs)), glutamate (GluCls) and histamine, as well as orthologues of the *Drosophila* pH-sensitive chloride channel[35]. The molecular diversity of the *Tribolium* Cys-loop LGIC superfamily is broadened by alternative splicing and RNA A-to-I editing, which in some cases generates species-specific receptor isoforms[35]. The *Tribolium* Cys-loop LGIC superfamily is the first complete set of genes encoding molecular targets of several insecticides—imidacloprid and other neonicotinoids (nAChRs), fipronil (GABARs) and avermectins (GluCls)—described for an agricultural pest species.

**Cytochrome P450 proteins.** Most insect cytochrome P450 proteins (CYPs) are thought to be involved in metabolic detoxification of host plant allelochemicals and toxicants, and several are insecticide resistance genes[36]. Other CYPs act in the synthesis and degradation of lipid signalling molecules, such as ecdysteroids[37]. Similarly to mosquitoes, especially *Aedes*, *Tribolium* has an independently expanded CYP gene family, particularly those involved in environmental response (Supplementary Table 16).

Within the *Tribolium* P450s, the CYP2 and mitochondrial clans have undergone relatively little gene expansion, lack pseudogenes, and are probably reserved for essential endogenous functions in ecdysteroid metabolism and development. In contrast, expansions via tandem duplication produced 85% of *Tribolium* P450s clustered in groups of 2–16 genes, with large expansions of CYP3 and CYP4 clans involved in environmental response. In comparison, *Apis* has only four CYP4 genes, whereas *Aedes* has relatively similarly sized expansions of CYP3 and CYP4 clans (Supplementary Table 16). We speculate that both mosquito larvae (which are omnivorous scavengers) and *Tribolium* have adapted to diverse chemical environments in part by expansion of CYP gene families involved in detoxification.

### Possible future insect control targets

**C1 cysteine peptidase genes.** *Tribolium castaneum* has successfully exploited cereal grains in spite of the arsenal of defensive allelochemicals, including inhibitors of serine peptidase digestive enzymes. In tenebrionid beetles, cathepsins B, L and serine peptidases such as trypsins and chymotrypsins are part of the digestive peptidase complex in the larval gut[38].

Comparing potential digestive peptidase genes in *Tribolium* with those in other sequenced insects (Supplementary Fig. 12) we found more C1 cysteine peptidase genes in *T. castaneum*. The proliferation of *Tribolium* C1 cysteine peptidase genes reflects expansions into five gene families, corresponding to four major clusters. This expansion is consistent with a trend seen in some beetles relative to other insects: a shift to a more acidic gut, conducive to cysteine peptidase activity.

*Tribolium castaneum* C1 cysteine peptidase genes encode B and L cathepsins, and include the first-known insect genes similar to O and K cathepsins (Supplementary Table 17). Most of the cathepsin-B-like peptidases lack conserved residues in functional regions and thus may lack peptidase activity, whereas all but two *Tribolium* cathepsin L peptidase genes encode potentially functional enzymes. In vertebrates, O and K cathepsins are lysosomal cysteine peptidases, involved in bone remodelling and resorption. Analysis of *Tribolium* cathepsins may provide insight into this family of proteins whose elevated expression is associated with a significant fraction of human breast cancers and tumour invasiveness.

**Neurohormones and G-protein-coupled receptors.** Insect neurohormones (neuropeptides, protein hormones and biogenic amines) control development, reproduction, behaviour, feeding and many other physiological processes, often by signalling through G-protein-coupled receptors (GPCRs). We found 20 genes encoding biogenic amine GPCRs in *Tribolium* (compared to 21 in *Drosophila* and 19 in *Apis*) and 52 genes encoding neuropeptide or protein hormone GPCRs (49 in *Drosophila*, 37 in *Apis*[39]). Moreover, we identified the likely ligands for 45 of these 72 *Tribolium* GPCRs. Furthermore, we annotated 39 neuropeptide and protein hormone genes. We found excellent agreement (95%) between the proposed ligands for the *Tribolium* neurohormone GPCRs and the independently annotated neuropeptide and protein hormone genes. Interestingly, the *Tribolium* genome contains a vasopressin-like neuropeptide (TC06626) and a vasopressin-like GPCR gene (TC16363; Supplementary Fig. 14), neither of which has been detected in any other sequenced insect[39]. Vasopressin in mammals is the major neurohormone stimulating water reabsorption in the kidneys[40]. Its presence in *Tribolium* may help the beetle to survive in very dry habitats.

## Genes relevant to *Tribolium* biology

**Vision.** Most of the 21 investigated genes that participate in the *Drosophila* photo-transduction network are conserved in *Tribolium* (Supplementary Table 14). Most notable is the lack of *ninaG* and *inaC*, which may be functionally replaced by closely related paralogues in *Tribolium*.

*Tribolium* contains only two opsin genes, representing members of the long-wavelength and ultraviolet-sensitivity-facilitating opsin subgroups. In contrast, *Drosophila* contains seven, and there is evidence for minimally three in most other insects. The lack of a blue-light-sensitive opsin gene in *Tribolium* is consistent with the unusual expression of long-wavelength opsin in all photoreceptor cells in this species[41]. The implied reduction in colour discrimination in *Tribolium* is probably a consequence of the widespread cryptic lifestyle of this species group.

**Odorant and gustatory receptors.** Odorant and gustatory receptors form the insect chemoreceptor superfamily. *Tribolium* has a major expansion of both odorant and gustatory receptors relative to *Drosophila*, *Anopheles* and *Aedes* mosquitoes, silkmoth and honeybee (Supplementary Table 19). We identified and annotated 265 apparently functional odorant receptors, 42 full-length pseudogenes and 34 pseudogene fragments. Most of these *T. castaneum* odorant receptors are in seven species-specific subfamilies, including one containing 150 genes, and most are in tandem gene arrays, created by gene duplication within the *Tribolium* lineage in the last 300 million years[42].

We annotated 220 apparently functional gustatory receptors and 25 pseudogenes (gustatory receptor gene fragments were not assessed). The gustatory receptor families in fruitflies and mosquitoes, but not honeybee, contain several genes that are alternatively spliced, with multiple alternative long first exons encoding at least the amino-terminal 50% of the gustatory receptor spliced into a set of short shared exons encoding the carboxy terminus[43–45]. Most *Tribolium* gustatory receptors are encoded by single genes; however, *T. castaneum* Gr214 is a massive alternatively spliced locus with 30 alternative long 5′ exons (six of which are pseudogenic) spliced into three shared 3′ exons encoding the C terminus. Three *T. castaneum* gustatory receptors are orthologues of highly conserved gustatory receptors in other insects[43–45], two of which form a heterodimeric carbon dioxide receptor[46]. The remainder form many species-specific subfamilies, one of which is expanded to 88 genes (Supplementary Information and Supplementary Fig. 16).

**Systemic RNAi.** In *Tribolium*, as in *C. elegans* but not *Drosophila*, the RNAi effect spreads systemically from the site of injection to other tissues[5] and from injected females to their offspring[4]. Surprisingly, our survey of genes involved in systemic RNA did not reveal much conservation between *Tribolium* and *C. elegans*.

The SID-1 multi-transmembrane protein, essential for double-stranded RNA (dsRNA) uptake in *C. elegans*, is not found in *Drosophila*, suggesting that the presence or absence of a *sid-1* gene is the primary determinant of whether or not systemic RNAi occurs in an organism. We found three genes in *Tribolium* that encode proteins similar to SID-1. However, their sequences are more similar to another *C. elegans* protein, TAG-130 (also known as ZK721.1), which is not required for systemic RNAi in *C. elegans*[47]. Additionally, the secondary argonaute proteins and RNA-dependent RNA polymerase (RdRP)[48,49], essential for the amplification of the initial dsRNA trigger in *C. elegans*, are absent in *Tribolium*. Therefore, the molecular basis for systemic RNAi in *Tribolium* and other insects might differ from that in *C. elegans* and remains to be elucidated.

## Concluding remarks

We observe three trends when comparing *Tribolium* and other insect genomes. First, phylogenetic trees show shorter branch lengths for *Tribolium* (and *Apis*) than *Drosophila*. The accelerated evolution of the *Drosophila* lineage in some cases rendered *Drosophila* atypical for the Insecta. Second, *Tribolium* retains a different set of ancestral genes that have evolved at a moderate rate (for example, *gremlin* and cathepsins), and these may provide insights into the function of their vertebrate orthologues. Third, its own evolutionary path has led to beetle- and perhaps *Tribolium*-specific gene changes (for example, a large increase in odorant receptors).

Expansions of CYP proteins, proteinases, diuretic hormones, a vasopressin hormone and receptor, and chemoreceptors all indicate adaptation to a dry, chemically diverse and toxin-rich microenvironment. Whereas the flour beetle's drought tolerance probably explains the presence of vasopressin, it is more difficult to rationalize a need for such an unprecedented diversity of chemoreceptors. Functions stemming from the diversity of angiosperm-derived chemicals such as distant detection of food sources and avoidance of toxic host plant defence chemicals suggest that this expansion may be common to the Coleoptera. The expansion of odorant receptors is more intriguing when considered in combination with the reduction of opsin genes. Both trends may reflect the long-term consequences of adaptation to low light biota by *Tribolium*, enforcing selection for increased discrimination of odour reception but not colour perception[41].

Given the chemo-sensing and detoxifying genes described above, it is perhaps no surprise that *Tribolium* has demonstrated resistance to all insecticides used for its control. Given the association of *Tribolium* with human food, knowledge of all possible insecticide targets will aid greater selectivity in pesticide design, thereby mitigating possible side effects. Finally, the true value of this sequence may be the entry it provides into the many and richly diverse facets of beetle biology, physiology and behaviour.

## METHODS SUMMARY

Detailed Methods are described in the Supplementary Information. Resources generated by this project can be found at the following locations: genome assemblies, sequences, and automated and manually curated gene model sequences are available from the BCM-HGSC website and ftp site (http://www.hgsc.bcm.tmc.edu/projects/tribolium/). Browser display of the genome sequence, all gene predictions and available tiling array data are available via http://www.genboree.org and Beetle Base (http://www.bioinformatics.ksu.edu/BeetleBase/), a long-term repository for *Tribolium* data.

**Received 16 July 2007; accepted 6 February 2008.**
**Published online 23 March 2008.**

1. Hunt, T. *et al.* A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318,** 1913–1916 (2007).
2. Sokoloff, A. *The Biology of* Tribolium *with Special Emphasis on Genetic Aspects I–III* (Clarendon Press and Oxford Univ. Press, Oxford, 1972, 1974, 1977).
3. Lorenzen, M. D. *et al. piggyBac*-based insertional mutagenesis in *Tribolium castaneum* using donor/helper hybrids. *Insect Mol. Biol.* **16,** 265–275 (2007).
4. Bucher, G., Scholten, J. & Klingler, M. Parental RNAi in *Tribolium* (Coleoptera). *Curr. Biol.* **12,** R85–R86 (2002).
5. Tomoyasu, Y. & Denell, R. E. Larval RNAi in *Tribolium* (Coleoptera) for analyzing adult development. *Dev. Genes Evol.* **214,** 575–578 (2004).
6. Tautz, D., Friedrich, M. & Schröder, R. in *Development 1994 Supplement* (eds Akam, M., Holland, P., Ingham, P. & Wray, G.) 193–199 (The Company of Biologists Limited, Cambridge, 1994).
7. Tautz, D. Segmentation. *Dev. Cell* **7,** 301–312 (2004).
8. Lorenzen, M. D. *et al.* Genetic linkage maps of the red flour beetle, *Tribolium castaneum*, based on bacterial artificial chromosomes and expressed sequence tags. *Genetics* **170,** 741–747 (2005).
9. Demuth, J. P. *et al.* Genome-wide survey of *Tribolium castaneum* microsatellites and description of 509 polymorphic markers. *Mol. Ecol. Notes* **7,** 1189–1195 (2007).
10. Wang, S. & Brown, S. J. Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome Biol.* (in the press).
11. Brown, S. J., Henry, J. K., Black, W. C. IV & Denell, R. Molecular Genetic manipulation of the red flour beetle: Genome organization and cloning of a ribosomal protein gene. *Insect Biochem.* **20,** 185–193 (1990).
12. Ugarkovic, D., Podnar, M. & Plohl, M. Satellite DNA of the red flour beetle *Tribolium castaneum*—comparative study of satellites from the genus *Tribolium. Mol. Biol. Evol.* **13,** 1059–1066 (1996).
13. Osanai, M., Kojima, K. K., Futahashi, R., Yaguchi, S. & Fujiwara, H. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). *Gene* **376,** 281–289 (2006).
14. Xiong, Y. & Eickbush, T. H. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. *Mol. Cell. Biol.* **8,** 114–123 (1988).
15. Robertson, H. M. & Gordon, K. H. Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera. Genome Res.* **16,** 1345–1351 (2006).
16. Fujiwara, H., Osanai, M., Matsumoto, T. & Kojima, K. K. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori. Chromosome Res.* **13,** 455–467 (2005).
17. Savard, J. *et al.* Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Res.* **16,** 1334–1338 (2006).
18. Savard, J., Tautz, D. & Lercher, M. J. Genome-wide acceleration of protein evolution in flies (Diptera). *BMC Evol. Biol.* **6,** 7 (2006).
19. Daborn, P. J. *et al.* A single p450 allele associated with insecticide resistance in *Drosophila. Science* **297,** 2253–2256 (2002).
20. Ciccarelli, F. D. *et al.* Complex genomic rearrangements lead to novel primate gene function. *Genome Res.* **15,** 343–351 (2005).
21. Wyder, S., Kriventseva, E. V., Schröder, R., Kadowaki, T. & Zdobnov, E. M. Quantification of ortholog losses in insects and vertebrates. *Genome Biol.* **8,** R242 (2007).
22. Wong, M. D., Jin, Z. & Xie, T. Molecular mechanisms of germline stem cell regulation. *Annu. Rev. Genet.* **39,** 173–195 (2005).

23. Schröder, R. The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium. Nature* **422,** 621–625 (2003).
24. Draper, B. W., Mello, C. C., Bowerman, B., Hardin, J. & Priess, J. R. MEX-3 is a KH domain protein that regulates blastomere identity in early *C. elegans* embryos. *Cell* **87,** 205–216 (1996).
25. Bucher, G. & Klingler, M. Divergent segmentation mechanism in the short germ insect *Tribolium* revealed by giant expression and function. *Development* **131,** 1729–1740 (2004).
26. Cerny, A. C., Bucher, G., Schröder, R. & Klingler, M. Breakdown of abdominal patterning in the *Tribolium Kruppel* mutant *jaws. Development* **132,** 5353–5363 (2005).
27. Choe, C. P., Miller, S. C. & Brown, S. J. A pair-rule gene circuit defines segments sequentially in the short-germ insect *Tribolium castaneum. Proc. Natl Acad. Sci. USA* **103,** 6560–6564 (2006).
28. Savard, J., Marques-Souza, H., Aranda, M. & Tautz, D. A segmentation gene in *Tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* **126,** 559–569 (2006).
29. Fonseca, R. N. *et al.* Self-regulatory circuits in dorsoventral axis formation of the short germ beetle *Tribolium castaneum*. *Dev. Cell* (in the press).
30. van der Zee, M., Stockhammer, O., von Levetzow, C., Nunes da Fonseca, R. & Roth, S. *Sog/Chordin* is required for ventral-to-dorsal *Dpp/BMP* transport and head formation in a short germ insect. *Proc. Natl Acad. Sci. USA* **103,** 16307–16312 (2006).
31. Beermann, A. & Schröder, R. Sites of FGF signalling and perception during embryogenesis of the beetle *Tribolium castaneum*. *Dev. Genes Evol.* (in the press).
32. Bolognesi, R. *et al. Tribolium* Wnts: evidence for a larger repertoire in insects with overlapping expression patterns that suggest multiple redundant functions in embryogenesis. *Dev. Genes Evol.* (in the press).
33. Beermann, A. *et al.* The *Short antennae* gene of *Tribolium* is required for limb development and encodes the orthologue of the *Drosophila* Distal-less protein. *Development* **128,** 287–297 (2001).
34. Raymond-Delpech, V., Matsuda, K., Sattelle, B. M., Rauh, J. J. & Sattelle, D. B. Ion channels: molecular targets of neuroactive insecticides. *Invert. Neurosci.* **5,** 119–133 (2005).
35. Jones, A. K. & Sattelle, D. B. The cys-loop ligand-gated ion channel gene superfamily of the red flour beetle, *Tribolium castaneum. BMC Genomics* **8,** 327 (2007).
36. Ffrench-Constant, R. H., Daborn, P. J. & Le Goff, G. The genetics and genomics of insecticide resistance. *Trends Genet.* **20,** 163–170 (2004).
37. Ono, H. *et al. Spook* and *Spookier* code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. *Dev. Biol.* **298,** 555–570 (2006).
38. Vinokurov, K. S. *et al.* Diversity of digestive proteinases in *Tenebrio molitor* (Coleoptera: Tenebrionidae) larvae. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **145,** 126–137 (2006).
39. Hauser, F., Cazzamali, G., Williamson, M., Blenau, W. & Grimmelikhuijzen, C. J. A review of neurohormone GPCRs present in the fruitfly *Drosophila melanogaster* and the honey bee *Apis mellifera. Prog. Neurobiol.* **80,** 1–19 (2006).
40. Bankir, L. Antidiuretic action of vasopressin: quantitative aspects and interaction between V1a and V2 receptor-mediated effects. *Cardiovasc. Res.* **51,** 372–390 (2001).
41. Jackowska, M. *et al.* Genomic and gene regulatory signatures of cryptozoic adaptation: loss of blue sensitive photoreceptors through expansion of long wavelength-opsin expression in the red flour beetle *Tribolium castaneum*. *Front. Zool.* **4,** 24 (2007).
42. Engsontia, P. *et al.* The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* (in the press).
43. Clyne, P. J., Warr, C. G. & Carlson, J. R. Candidate taste receptors in *Drosophila. Science* **287,** 1830–1834 (2000).
44. Hill, C. A. *et al.* G protein-coupled receptors in *Anopheles gambiae. Science* **298,** 176–178 (2002).
45. Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA* **100** (suppl. 2), 14537–14542 (2003).
46. Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila. Nature* **445,** 86–90 (2007).
47. Tomoyasu, Y. *et al.* Exploring systemic RNA interference in insects; a genome-wide survey for RNAi genes in *Tribolium. Genome Biol.* (in the press).
48. Lipardi, C., Wei, Q. & Paterson, B. M. RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* **107,** 297–307 (2001).
49. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107,** 465–476 (2001).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** Work at the BCM-HGSC was funded by grants from the NHGRI and USDA. FgenesH and FgenesH++ analysis was donated by Softberry Inc. This research was additionally supported in part by the Intramural Research Program of the NIH, National Library of Medicine.

**Author Information** The *Tribolium* genome sequence, at the NCBI, has project accession AAJJ00000000. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.R. (stephenr@bcm.edu).

**The *Tribolium* Genome Sequencing Consortium**

**Project leader:** Stephen Richards[1,2]

**Principal investigators:** Richard A. Gibbs[1,2], George M. Weinstock[1,2]

**White paper:** Susan J. Brown[3], Robin Denell[3], Richard W. Beeman[4], Richard Gibbs[1,2]

**Analysis leaders:** Richard W. Beeman[4], Susan J. Brown[3], Gregor Bucher[5], Markus Friedrich[6] Cornelis J. P. Grimmelikhuijzen[7], Martin Klingler[8], Marce Lorenzen[3], Stephen Richards[1,2], Siegfried Roth[9], Reinhard Schröder[10,†], Diethard Tautz[11], Evgeny M. Zdobnov[12,13,14]

**DNA sequence and global analysis: DNA sequencing** Donna Muzny (leader)[1,2], Richard A. Gibbs[1,2], George M. Weinstock[1,2], Tony Attaway[1,2], Stephanie Bell[1,2], Christian J. Buhay[1,2], Mimi N. Chandrabose[1,2], Dean Chavez[1,2], Kerstin P. Clerk-Blankenburg[1,2], Andrew Cree[1,2], Marvin Dao[1,2], Clay Davis[1,2], Joseph Chacko[1,2], Huyen Dinh[1,2], Shannon Dugan-Rocha[1,2], Gerald Fowler[1,2], Toni T. Garner[1,2], Jeffrey Garnes[16], Andreas Gnirke[17], Alica Hawes[1,2], Judith Hernandez[1,2], Sandra Hines[1,2], Michael Holder[1,2], Jennifer Hume[1,2], Shalini N. Jhangiani[1,2], Vandita Joshi[1,2], Ziad Mohid Khan[1,2], LaRonda Jackson[1,2], Christie Kovar[1,2], Andrea Kowis[1,2], Sandra Lee[1,2], Lora R. Lewis[1,2], Jon Margolis[17], Margaret Morgan[1,2], Lynne V. Nazareth (leader)[1,2], Ngoc Nguyen[1,2], Geoffrey Okwuonu[1,2], David Parker[1,2], Stephen Richards[1,2], San-Juana Ruiz[1,2], Jireh Santibanez[1,2], Joël Savard[11], Steven E. Scherer[1,2], Brian Schneider[1,2], Erica Sodergren[1,2], Diethard Tautz[11], Selina Vattahil[1,2], Donna Villasana[1,2], Courtney S. White[1,2], Rita Wright[1,2]; **EST sequencing** Yoonseong Park[18], Richard W. Beeman[4], Jeff Lord[4], Brenda Oppert[4], Marce Lorenzen[4], Susan Brown[3], Liangjiang Wang[3], Joël Savard[11], Diethard Tautz[11], Stephen Richards[1], George Weinstock[1,2], Richard A. Gibbs[1,2]; **genome assembly** Yue Liu[1,2], Kim Worley[1,2], George Weinstock[1,2]; **G+C content** Christine G. Elsik[19], Justin T. Reese[19], Eran Elhaik[20], Giddy Landan[20], Dan Graur[20]; **repetitive DNA, transposons and telomeres** Peter Arensburger[21], Peter Atkinson[21], Richard W. Beeman[4], Jim Beidler[22], Susan J. Brown[3], Jeffery P. Demuth[23], Douglas W. Drury[24], Yu-Zhou Du[25], Haruhiko Fujiwara[26], Marce Lorenzen[3], Vincenza Maselli[27], Mizuko Osanai[26], Yoonseong Park[18], Hugh M. Robertson[28], Zhijian Tu[22], Jian-jun Wang[25], Suzhi Wang[3]; **gene prediction and consensus gene set** Stephen Richards[1,2], Henry Song[1,2], Lan Zhang[1,2], Erica Sodergren[1,2], Doreen Werner[29], Mario Stanke[29], Burkhard Morgenstern[29], Victor Solovyev[30], Peter Kosarev[31], Garth Brown[32], Hsiu-Chuan Chen[32], Olga Ermolaeva[32], Wratko Hlavina[32], Yuri Kapustin[32] Boris Kiryutin[32], Paul Kitts[32], Donna Maglott[32], Kim Pruitt[32], Victor Sapojnikov[32], Alexandre Souvorov[32], Aaron J. Mackey[33], Robert M. Waterhouse[14], Stefan Wyder[12], Evgeny M. Zdobnov[12,13,14]; **global gene content analysis** Evgeny M. Zdobnov[12,13,14], Stefan Wyder[12], Evgenia V. Kriventseva[12,34], Tatsuhiko Kadowaki[35], Peer Bork[36,37]

**Developmental processes and signalling pathways:** Manuel Aranda[11], Riyue Bao[6], Anke Beermann[10], Nicola Berns[10], Renata Bolognesi[3], François Bonneton[38], Daniel Bopp[39], Susan J. Brown[3], Gregor Bucher[5], Thomas Butts[40], Arnaud Chaumot[41], Robin E. Denell[3], David E. K. Ferrier[40], Markus Friedrich[6], Cassondra M. Gordon[3], Marek Jindra[42], Martin Klingler[8], Que Lan[43], H. Michael G. Lattorff[44], Vincent Laudet[38], Cornelia von Levetsow[9], Zhenyi Liu[45], Rebekka Lutz[10], Jeremy A. Lynch[9], Rodrigo Nunes da Fonseca[9], Nico Posnien[5], Rolf Reuter[10], Siegfried Roth[9], Joël Savard[11], Johannes B. Schinko[5], Christian Schmitt[8], Michael Schoppmeier[8], Reinhard Schröder[10], Teresa D. Shippy[3], Franck Simonnet[5], Henrique Marques-Souza[11], Diethard Tautz[11], Yoshinori Tomoyasu[3], Jochen Trauner[8], Maurijn Van der Zee[11], Michel Vervoort[46], Nadine Wittkopp[10], Ernst A. Wimmer[5], Xiaoyun Yang[6]

**Pest biology, senses, Medea and RNAi: ligand gated ion channels** Andrew K. Jones[47], David B. Sattelle[47]; **oxidative phosphorylation** Paul R. Ebert[48]; **P450 genes** David Nelson[49], Jeffrey G. Scott[50], Richard W. Beeman[4]; **chitin and cuticular proteins** Subbaratnam Muthukrishnan[51], Karl J. Kramer[4,51], Yasuyuki Arakane[4,51], Richard W. Beeman[4], Qingsong Zhu[51], David Hogenkamp[51], Radhika Dixit[51]; **digestive proteinases** Brenda Oppert[4], Haobo Jiang[52], Zhen Zou[52], Jeremy Marshall[3], Elena Elpidina[53], Konstantin Vinokurov[53], Cris Oppert[4]; **immunity** Zhen Zou[52], Jay Evans[54], Zhiqiang Lu[52], Picheng Zhao[52], Niranji Sumathipala[52], Boran Altincicek[55], Andreas Vilcinskas[55], Michael Williams[56], Dan Hultmark[56], Charles Hetru[57], Haobo Jiang[52]; **neurohormones and GPCRs** Cornelis J. P. Grimmelikhuijzen[7], Frank Hauser[7], Giuseppe Cazzamali[7], Michael Williamson[7], Yoonseong Park[18], Bin Li[18], Yoshiaki Tanaka[58], Reinhard Predel[59], Susanne Neupert[59], Joachim Schachtner[60], Peter Verleyen[61]; **neuropeptide processing enzymes** Florian Raible[36], Peer Bork[36,37]; **opsins** Markus Friedrich[6]; **odorant receptors and gustatory receptors** Kimberly K. O. Walden[28], Hugh M. Robertson[28]; **odorant binding and chemosensory proteins** Sergio Angeli[62], Sylvain Forêt[63], Gregor Bucher[5], Stefan Schuetz[5], Ryszard Maleszka[63], Ernst A. Wimmer[5]; **Medea** Richard W. Beeman[4], Marce Lorenzen[4]; **systemic RNAi** Yoshinori Tomoyasu[3], Sherry C. Miller[3], Daniela Grossmann[5] & Gregor Bucher[5]

Affiliations for participants: [1]Human Genome Sequencing Center, and [2]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [3]Division of Biology, Ackert Hall, Kansas State University, Manhattan, Kansas 66506, USA. [4]Grain Marketing and Production Research Center, Agricultural Research Service, United States Department of Agriculture, 1515 College Avenue, Manhattan, Kansas 66502, USA. [5]Johann Friedrich Blumenbach Institute, Department of Developmental Biology, Georg August University, von-Liebig-Weg-11, 37077 Göttingen, Germany. [6]Department of Biological Sciences, Wayne State University, Detroit, Michigan 48202, USA. [7]Center for Functional and Comparative Insect Genomics, and Department of Cell Biology and Comparative Zoology, Institute of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. [8]Institute for Biology, Department of Developmental Biology, Friedrich-Alexander-University Erlangen, Staudtstrasse 5, 91058 Erlangen, Germany. [9]Institute for Developmental Biology, University of Cologne, 50674 Cologne, Germany. [10]Animal Genetics, Interfaculty Institute for Cell Biology, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany. [11]Department of Genetics, University of Cologne, 50674 Cologne, Germany. [12]Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [13]Swiss Institute of Bioinformatics, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [14]Imperial College London, South Kensington Campus, SW7 2AZ London, UK. [15]Children's Hospital Oakland Research Institute, BACPAC Resources, 747 52nd Street, Oakland, California 94609, USA. [16]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [17]AgraQuest, Inc., 1530 Drew Avenue, Davis, California 95616, USA. [18]Department of Entomology, Waters Hall, Kansas State University, Manhattan, Kansas 66506, USA. [19]Department of Animal Science, Texas A&M University, College Station, Texas 77843, USA. [20]Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, USA. [21]Department of Entomology, University of California, 900 University Avenue, Riverside, California 92521, USA. [22]Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061, USA. [23]Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA. [24]Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. [25]Department of Plant Protection, Yangzhou University, Yangzhou 225009, China. [26]Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Bioscience Building 501, Kashiwa, Chiba 277-8562, Japan. [27]European School of Molecular Medicine and Telethon Institute of Genetics and Medicine, Via Pietro Castellino 111, 80131 Napoli, Italy. [28]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [29]Institute for Microbiology and Genetics, Department of Bioinformatics, University of Göttingen, Goldschmidtstraße 1, 37077 Göttingen, Germany. [30]Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. [31]Softberry Inc., 116 Radio Circle, Suite 400, Mount Kisco, New York 10549, USA. [32]National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland 20894, USA. [33]GlaxoSmithKline, Collegeville, Pennsylvania 19426, USA. [34]Department of Structural Biology and Bioinformatics, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [35]Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan. [36]European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany. [37]Max-Delbruck-Centre for Molecular Medicine, Berlin-Buch, Robert-Roessle-Strasse 10, 13092 Berlin, Germany. [38]Institut de Genomique Fonctionnelle de Lyon, Equipe de Zoologie Moleculaire, ENS Lyon, Universite Lyon 1, CNRS UMR5242, INRA, IFR128, 46 Allee d'Italie, 69364 Lyon cedex 07, France. [39]Zoological Institute of the University Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [40]Department of Zoology, University of Oxford, Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK. [41]CEMAGREF, Laboratoire d'écotoxicologie, 3bis quai Chauvea, CP220 69336 Lyon cedex 09, France. [42]Institute of Entomology ASCR, Branisovká 31, Ceské Budejovice 370 05, Czech Republic. [43]Department of Entomology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, Wisconsin 53706, USA. [44]Institute of Biology, Molecular Ecology, Martin-Luther-University Halle-Wittenberg Hoher Weg 4, 06099 Halle (Saale), Germany. [45]Department of Molecular Biology and Pharmacology, Washington University in St Louis School of Medicine, 3600 Cancer Research Building, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [46]Université Paris 7 – Denis Diderot, Centre de genetique moleculaire – CNRS UPR 2167, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette cedex, France. [47]MRC Functional Genetics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK. [48]School of Integrative Biology & School of Molecular and Microbial Sciences, University of Queensland, St Lucia, Queensland 4072, Australia. [49]Department of Molecular Sciences and Center of Excellence in Genomics and Bioinformatics, University of Tennessee, Memphis, Tennessee 38163, USA. [50]Department of Entomology, Daljit and Elaine Sarkaria Professor of Insect Physiology and Toxicology, Cornell University, Ithaca, New York 14853, USA. [51]Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506, USA. [52]Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma 74078, USA. [53]A. N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Leninskie Gory, Moscow 119992, Russia. [54]USDA-ARS Bee Research Laboratory, Beltsville, Maryland 20705, USA. [55]Institute of Phytopathology and Applied Zoology, Interdisciplinary Research Center, Justus-Liebig-University of Giessen, Heinrich-Buff-Ring 26-32, D-35392 Giessen, Germany. [56]Umea Centre for Molecular Pathogenesis, Umea University, Umea SE-90187, Sweden. [57]Institut Biol Moléc Cell, CNRS, Strasbourg 67084, France. [58]National Institute of Agrobiological Science, Division of Insect Science, Tsukuba, Ibaraki 305-8634, Japan. [59]Institute of General Zoology, University of Jena, Erbertstrasse 1, D-07743 Jena, Germany. [60]Department of Animal Physiology, University of Marburg, Karl-von-Frisch Strasse 8, D-35032 Marburg, Germany. [61]Department of Animal Physiology and Neurobiology, University of Leuven, Naamsestraat 59, BE-3000 Leuven, Belgium. [62]Institute for Forest Zoology and Forest Conservation, Büsgenweg 3 D-37077 Göttingen, Germany. [63]Visual Sciences and ARC Centre for the Molecular Genetics of Development, Research School of Biological Sciences, The Australian National University, Canberra, ACT 0200, Australia. †Present address: Bioscience Institute, University of Rostock, Albert-Einstein-Strasse 3, 18059 Rostock, Germany.

Vol 452|24 April 2008|doi:10.1038/nature06864

# ARTICLES

# Lateral presynaptic inhibition mediates gain control in an olfactory circuit

Shawn R. Olsen[1] & Rachel I. Wilson[1]

**Olfactory signals are transduced by a large family of odorant receptor proteins, each of which corresponds to a unique glomerulus in the first olfactory relay of the brain. Crosstalk between glomeruli has been proposed to be important in olfactory processing, but it is not clear how these interactions shape the odour responses of second-order neurons. In the *Drosophila* antennal lobe (a region analogous to the vertebrate olfactory bulb), we selectively removed most interglomerular input to genetically identified second-order olfactory neurons. Here we show that this broadens the odour tuning of these neurons, implying that interglomerular inhibition dominates over interglomerular excitation. The strength of this inhibitory signal scales with total feedforward input to the entire antennal lobe, and has similar tuning in different glomeruli. A substantial portion of this interglomerular inhibition acts at a presynaptic locus, and our results imply that this is mediated by both ionotropic and metabotropic receptors on the same nerve terminal.**

A sensory stimulus generally triggers activity in multiple neural processing channels, each of which carries information about some feature of that stimulus. The concept of a processing channel has a particularly clear anatomical basis in the first relay of the olfactory system, which is typically divided into glomerular compartments. Each glomerulus receives input from many first-order olfactory receptor neurons (ORNs), all of which express the same odorant receptor. Each second-order neuron receives direct ORN input from a single glomerulus, and thus all the first- and second-order neurons corresponding to a glomerulus constitute a discrete processing channel. An odorant typically triggers activity in multiple glomeruli, and local interneurons that interconnect glomeruli provide a substrate for crosstalk between channels.

The *Drosophila* antennal lobe is a favoured model for investigating olfactory processing because it contains only ~50 glomeruli[1], each of which corresponds to an identified type of ORN and an identified type of postsynaptic projection neuron (PN)[2–5]. Several recent studies of the *Drosophila* antennal lobe have produced divergent views regarding the relative importance of interglomerular connections. One model proposes that PN odour responses are almost completely determined by feedforward excitation[6,7]. This model ascribes little importance to crosstalk between glomerular processing channels. An alternative model proposes that interglomerular connections make an important contribution to shaping PN odour responses[8–12]. However, this has not been demonstrated by showing a change in PN odour responses when lateral inputs to that PN are removed. (We use the word 'lateral' as a synonym for 'interglomerular'.)

In principle, several features of olfactory processing in the *Drosophila* antennal lobe could reflect either intra- or interglomerular events. For example, most PNs are more broadly tuned to odours than their presynaptic ORNs[8,10]. This could reflect a purely intraglomerular nonlinear process, such as short-term synaptic depression at ORN–PN connections. Alternatively, it could be due to the fact that lateral excitatory connections exist between glomeruli[7,11,12]. It is also unclear whether inhibitory epochs in PN odour responses[8–10] reflect inter- or intraglomerular events. Many interneurons that are immuno-positive for GABA ($\gamma$-aminobutyric acid) form connections between glomeruli[9,13,14], but several recent studies have failed to observe any interglomerular inhibition[7,11,12]. These studies removed all the direct ORN inputs to an identified PN, and asked whether lateral input could be evoked in that PN by olfactory stimulation of other ORN types. In all cases, lateral inputs to PNs were excitatory. This raises the possibility that interglomerular inhibition might not exist, and that inhibitory PN odour responses might merely reflect intraglomerular feedback[15–17]. This is an important issue because intra- and interglomerular inhibition have different consequences for how olfactory representations are transformed in this circuit.

In this study, we address three questions. Do interglomerular interactions make a substantial contribution to PN odour responses? Do these interactions include lateral inhibition? If so, how does this occur, and why has it been difficult to observe?

## Removing lateral input to projection neurons

We began by investigating what happens to PN odour responses when most lateral input to that PN is removed. We took advantage of the fact that the fly has two olfactory organs. About 90% of ORNs are contained in the antennae, whereas 10% are in the maxillary palp (Fig. 1a). Palp ORNs express odorant receptors not expressed in the antennae, and project to palp glomeruli that are distinct from glomeruli targeted by antennal ORNs[3,4]. Antennal and palp glomeruli are interconnected by local interneurons. Acute removal of the antennae eliminates 90% of the input to the antennal lobe, and therefore most excitatory drive to local interneurons (Fig. 1b). If lateral connections are mainly excitatory[7,11,12], then removing the antennae should decrease the odour responses of palp PNs.

We performed this experiment in two different palp glomeruli (VM7 and VC1). Surprisingly, removing the antennae increased most of the odour responses of these PNs (Fig. 1c, d and Supplementary Figs 2 and 3). No odour responses were decreased. This implies that most of our odours normally evoke lateral inhibitory input to these glomeruli, and that this outweighs the effect of lateral excitatory input.

We examined the input–output function of each glomerulus by plotting the strength of each PN odour response against the strength of the cognate ORN response to the same odour. These input–output functions were nonlinear (Fig. 1d), which is typical for most glomeruli[10]. When we removed most lateral input to VM7 and VC1 PNs,

[1]Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.

the nonlinearity persisted (Fig. 1d) and these PNs became even more broadly tuned (Fig. 1e). This argues that broad tuning of PNs[8,10] results mainly from purely intraglomerular mechanisms. These could include short-term synaptic depression at ORN–PN connections and/or an intrinsic ceiling on PN firing rates. Lateral excitation should tend to broaden PN tuning even more, but lateral inhibition evidently counteracts this.

One clue to the significance of lateral inputs is that, in the intact antennal lobe circuit, PN responses cannot be predicted purely on the basis of feedforward excitatory inputs[10]. Two odours can elicit similar responses in an ORN but divergent responses in a postsynaptic PN. For example, pentyl acetate and 4-methyl phenol evoke similar activity in VM7 ORNs but not in VM7 PNs (Fig. 1c), implying that these odours recruit different lateral inputs to this glomerulus. After antennae were removed, these odours evoked similar responses in VM7 PNs (Fig. 1c). Overall, removing the antenna increased the correlation between the ranked odour preferences of ORNs and their cognate PNs (Spearman's $\rho$ increased from 0.82 to 0.88 for VM7 and from 0.89 to 0.94 for VC1; $P < 0.02$ for each comparison, Mann–Whitney *U*-test).

## Lateral inhibition suppresses ORN input

Several recent studies have failed to observe lateral inhibition in the *Drosophila* antennal lobe[7,11,12]. These studies silenced all direct ORN inputs to a PN and focused on lateral input to that PN evoked by stimulating ORNs presynaptic to other glomeruli. We reasoned that some lateral inhibition might target ORN axon terminals; if so, this would only be observed when the direct ORN inputs to a PN are active. GABAergic inhibition at ORN axon terminals in the olfactory bulb has been described previously[15–21], and synapses from GABAergic interneurons onto ORN axon terminals have been found in an insect antennal lobe[22].

To test the idea that some lateral inhibition is presynaptic, we asked how lateral input to glomerulus VM7 depends on the activity of VM7 ORNs. Each VM7 ORN fires spontaneously at ~10 spikes s$^{-1}$, and consequently these PNs are bombarded by spontaneous spike-driven excitatory postsynaptic potentials (EPSPs; Supplementary Fig. 4a). When we silenced VM7 ORNs by removing the maxillary palps, large spontaneous EPSPs disappeared in VM7 PNs (Fig. 2a). In this experimental configuration, stimulating the antennae with odorants depolarized VM7 PNs (Fig. 2a), which is consistent with previous reports that lateral input is excitatory when direct ORN inputs are silent[7,11,12]. Next, we asked whether preserving spontaneous activity in VM7 ORNs would allow us to observe lateral inhibition. To prevent odour-evoked activity in VM7 ORNs, we covered the maxillary palps

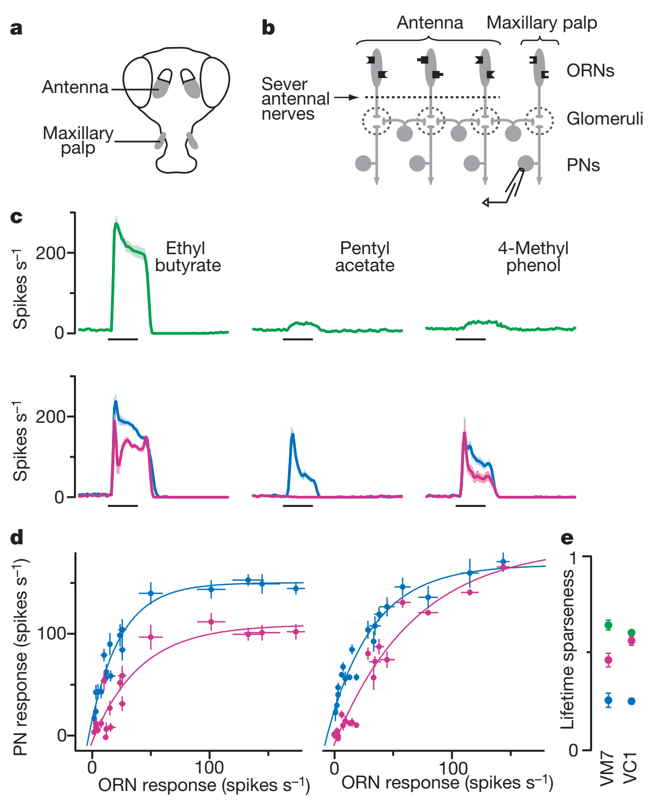



**Figure 1 | Removing lateral input disinhibits projection neurons.** **a**, *Drosophila* olfactory organs. **b**, Experimental configuration. **c**, Spiking responses of ORNs (green) and PNs (blue and magenta) for the palp glomerulus VM7 ($n = 4$–12 for each response). PNs are shown with (magenta) and without (blue) lateral input from antennal glomeruli. The black bars represent the 500-msec odour-stimulus period. **d**, Input–output functions for palp glomeruli VM7 (left) and VC1 (right). Each point represents the average PN response to an odour versus the response of the cognate ORNs. PNs are shown with antennae intact (magenta) or removed (blue). Responses to 18 out of 20 odours are significantly disinhibited in VM7; 13 out of 20 are significantly disinhibited in VC1 ($P < 0.05$, *t*-tests). **e**, Odour selectivity of ORNs (green) and cognate PNs with antennae intact (magenta) or removed (blue). Lifetime sparseness is 0 for an unselective cell and 1 for a maximally selective cell. All within-glomerulus comparisons are significant ($P < 0.002$, Mann–Whitney *U*-test), except ORNs versus antennae-intact PNs for VC1. Data are shown as mean ± s.e.m.

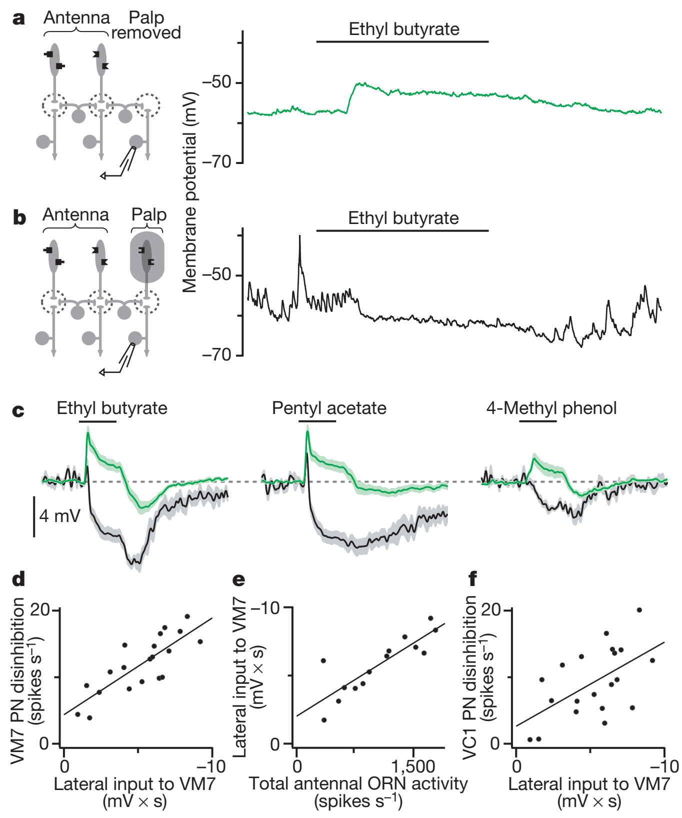



**Figure 2 | Lateral inhibition suppresses spontaneous EPSCs and scales with total ORN input.** **a**, Left: experimental configuration. Right: recording from a PN in glomerulus VM7. Olfactory stimulation evokes depolarization. Spontaneous EPSPs are absent. **b**, Left: palps are shielded from odours. Right: recording from a PN in glomerulus VM7. Olfactory stimulation suppresses spontaneous EPSPs. **c**, Average VM7 PN responses as in **a** (green) or **b** (black); $n = 6$–7 PNs for each condition. When ORNs are shielded (black), inhibition dominates. When ORNs are absent (green), excitation dominates throughout the stimulus period. (Off-inhibition probably reflects lateral postsynaptic inhibition, see Supplementary Fig. 5.) Pale bands are s.e.m. **d**, Lateral input to VM7 (as in **b**) is correlated with the disinhibition in VM7 PNs after antennal removal (see Fig. 1b). Each point represents a different odour. Lateral input is measured as the time-integrated change in membrane potential. **e**, Lateral input to VM7 is correlated with total antennal ORN spiking activity evoked by each odour. **f**, Lateral input to VM7 is correlated with the disinhibition in VC1 PNs after antennal removal.

with a plastic shield. We inferred that the shield did not prevent spontaneous activity in VM7 ORNs because we observed normal spontaneous EPSPs in VM7 PNs (Fig. 2b). The shield was clearly an effective barrier to odours because it blocked the normal strong excitatory response to ethyl butyrate in VM7 PNs (Supplementary Fig. 4). Instead, when the palps were shielded, stimulating the antennae with odorants suppressed spontaneous EPSPs in VM7 PNs (Fig. 2b, c). This was accompanied by hyperpolarization of the membrane potential, which would reflect (at least in part) the removal of ongoing depolarization produced by EPSP bombardment.

For all 20 odours in our panel, lateral input depolarized VM7 PNs when their cognate ORNs were absent (as in Fig. 2a) and hyperpolarized these PNs when their ORNs were spontaneously active (as in Fig. 2b). This argues that a substantial component of lateral inhibition acts at a presynaptic locus. This does not mean that all lateral inhibition is presynaptic; indeed there is evidence for an additional postsynaptic component (Supplementary Fig. 5).

### Inhibition scales with total ORN input

When palp ORNs were shielded, different odours evoked different amounts of inhibition in glomerulus VM7 (Fig. 2c, black traces). We hypothesized that this odour tuning could explain why antennal removal disinhibits some VM7 PN odour responses more than others (Fig. 1c, d). To test this, we measured the amount of inhibition

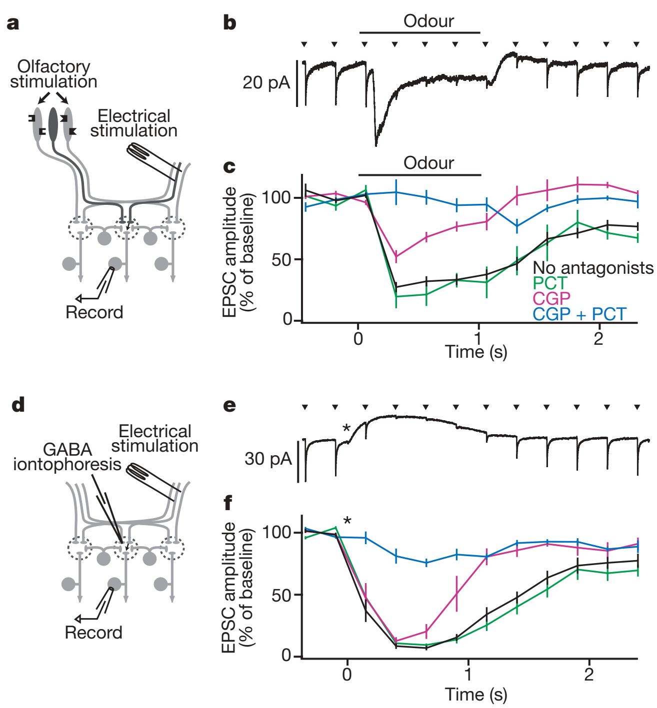



**Figure 3 | Lateral GABAergic suppression of ORN–PN synapses.** **a**, Experimental configuration. **b**, Electrical stimulation of the antennal nerve (arrowheads) evokes EPSCs in a PN (average of 20 trials). Olfactory stimulation (500 ms) inhibits EPSCs. Odour also evokes a transient inward current reflecting lateral postsynaptic excitation; this is resistant to GABA antagonists (Supplementary Fig. 6). **c**, Inhibition is blocked by the $GABA_B$ antagonist CGP54626 (CGP) together with the $GABA_A$ antagonist picrotoxin (control, $n = 12$; PCT, $n = 5$; CGP, $n = 5$; CGP + PCT, $n = 5$). All pairwise comparisons are significantly different except control versus PCT ($P < 0.05$, $t$-tests). **d**, Experimental configuration, substituting GABA iontophoresis for olfactory stimulation. **e**, As in **b** for GABA iontophoresis (asterisk). GABA also evokes an outward current. **f**, As in **c** for GABA iontophoresis (control, $n = 13$; PCT, $n = 5$; CGP, $n = 6$; CGP + PCT, $n = 7$). All pairwise comparisons between conditions are significantly different except control versus PCT ($P < 0.05$, $t$-tests). Data are shown as mean ± s.e.m.

evoked by each odour in the shielded-palps experiment, and compared this to the change in PN spiking responses to that odour after antennal removal. These two measures were well correlated (Fig. 2d, $n = 20$ odours, Pearson's $r^2 = 0.65$, $P < 0.0001$), which argues that these two experimental paradigms measure the same underlying phenomenon.

The odour tuning of lateral input must reflect the connectivity of the local interneurons that mediate this inhibition. Many individual GABAergic interneurons innervate all glomeruli[9,12,13], suggesting that lateral inhibitory input to each glomerulus might reflect pooled input from all ORNs. If so, then the size of lateral input to a glomerulus should correlate with the total ORN activity evoked by that odour. We estimated total ORN activity by summing the spiking responses of each antennal ORN type[23], and found that this measure predicted the strength of lateral inhibition evoked by each odour in the shielded-palps experiment (Fig. 2e, $n = 14$ odours, Pearson's $r^2 = 0.73$, $P < 0.0005$).

If lateral inhibition to all glomeruli scales with total ORN activity, then lateral inhibitory input to each glomerulus would show the same odour tuning. We have already shown that the odour tuning of lateral input to VM7 is a good predictor of which odour responses were most disinhibited in VM7 PNs after antennal removal (Fig. 2d). As expected, it also partially predicted which odour responses were most disinhibited in a different palp glomerulus, VC1 (Fig. 2f, $n = 20$ odours, Pearson's $r^2 = 0.32$, $P < 0.01$). However, this correlation was weaker than the correlation with disinhibition in VM7. This leaves open the possibility that there may be some differences in the odour tuning of lateral inhibitory input to different glomeruli (see Discussion).

### GABA mediates presynaptic inhibition

Our results suggest that much of the lateral inhibition in this circuit acts by suppressing ORN–PN synaptic transmission. To test this, we monitored ORN–PN synaptic strength in one glomerulus while recruiting lateral input to that glomerulus (Fig. 3a). We recorded from an identified PN while electrically stimulating the ipsilateral antennal nerve to evoke excitatory postsynaptic currents (EPSCs). Next, we used an odour to stimulate ORNs in the remaining intact antenna (and the maxillary palps). Because most glomeruli receive bilateral ORN input[2], olfactory stimulation of the contralateral antenna drives activity in ipsilateral glomeruli. Finally, we prevented odours from recruiting direct ORN input to the recorded PN by mutating the odorant receptor gene normally expressed by its ORNs.

As predicted, olfactory stimulation of the contralateral antenna inhibited EPSCs evoked by ipsilateral nerve stimulation (Fig. 3b, c). We could mimic this inhibition by iontophoresing GABA into the antennal lobe neuropil (Fig. 3d–f). A $GABA_B$ receptor antagonist blocked the late phase of this inhibition, but had only a modest effect on the early phase (Fig. 3c, f). Adding a $GABA_A$ antagonist to the $GABA_B$ antagonist blocked the residual early portion of the inhibition (Fig. 3c, f). The $GABA_A$ antagonist alone had no effect (Fig. 3c, f). Taken together, these results suggest that both $GABA_A$ and $GABA_B$ receptors are present on the same ORN axon terminals, and that either $GABA_A$ or $GABA_B$ receptors alone are sufficient to mediate substantial inhibition of EPSCs just after GABA release. The late phase of inhibition evidently involves only $GABA_B$ receptors.

Presynaptic inhibition is generally associated with a change in the way in which a synapse responds to paired electrical pulses[24]. We found that both the $GABA_A$ and the $GABA_B$ components of EPSC inhibition are associated with an increase in the paired-pulse ratio (Supplementary Fig. 7). This implies that the independent actions of both $GABA_A$ and $GABA_B$ receptors are at least partially presynaptic. As a further test of this model, we genetically abolished $GABA_B$ signalling selectively in presynaptic ORNs. We used an ORN-specific promoter[25] to drive expression of pertussis toxin[26]—a selective inhibitor of some types of G proteins. In these flies, GABA still inhibited ORN–PN EPSCs, but now this inhibition had a briefer duration than

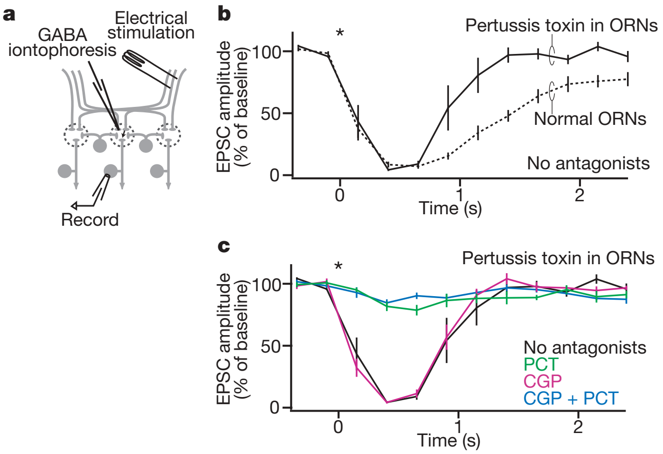



**Figure 4 | Genetic evidence that $GABA_B$ receptors inhibit ORN–PN synapses at a presynaptic locus.** **a**, Experimental configuration. **b**, When pertussis toxin is specifically expressed in ORNs, GABAergic inhibition of EPSCs (solid line, $n = 11$) is more transient than in control flies (dotted line, $n = 13$, reproduced from Fig. 3f). **c**, Pertussis toxin expression in ORNs renders the GABAergic inhibition of EPSCs completely insensitive to CGP, and completely sensitive to PCT (black trace reproduced from **a**; CGP, $n = 5$; PCT, $n = 6$; CGP + PCT, $n = 5$). Compare to Fig. 3f. Asterisk represents GABA iontophoresis. Data are shown as mean ± s.e.m.

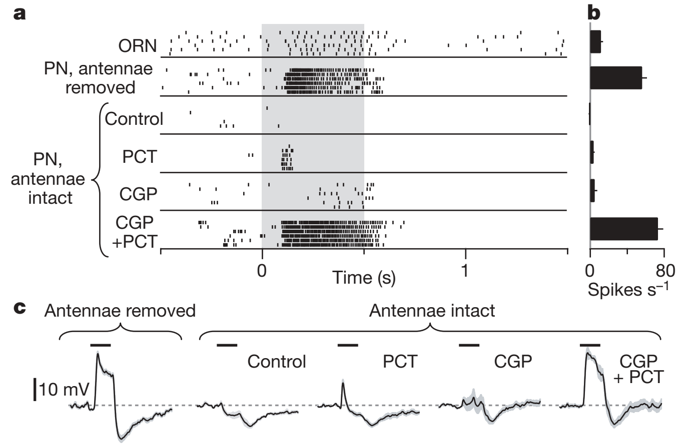



**Figure 5 | GABA receptor antagonists mimic removal of lateral input to a projection neuron.** **a**, Rasters show spiking responses to pentyl acetate (grey) in a VM7 ORN and a VM7 PN. With antennae intact, both antagonists are required to mimic the effect of antennal removal on PNs. (Some off-inhibition in PNs persists after antennal removal or in the presence of antagonists; this probably reflects the off-inhibition in VM7 ORNs.) **b**, Average spike rates during odour stimulus period, minus baseline spike rates ($n = 5$–6 PNs for each condition). **c**, Average membrane potential responses to pentyl acetate in VM7 PNs ($n = 5$–6 projection neurons for each). Data are shown as mean ± s.e.m.

in wild-type flies (Fig. 4a, b). Unlike in wild-type flies, this inhibition was completely resistant to the $GABA_B$ antagonist and completely blocked by the $GABA_A$ antagonist (compare Fig. 4c to Fig. 3f). As a negative control, we confirmed that this phenotype requires both the ORN-specific promoter and the toxin transgene (Supplementary Fig. 8). This demonstrates that $GABA_B$ receptors inhibit ORN–PN synapses at a purely presynaptic locus.

If activation of either $GABA_A$ or $GABA_B$ receptors is sufficient to mediate strong lateral inhibition, then blockade of both receptors should be required to mimic the removal of lateral input. To test this, we again recorded from palp PNs in glomerulus VM7. Normally, the odour pentyl acetate weakly excites VM7 ORNs and inhibits VM7 PNs. When most lateral input is removed (by removing the antennae), this odour strongly excites these PNs (Fig. 1c). We could not mimic this disinhibition by applying either a $GABA_A$ or a $GABA_B$ receptor antagonist alone. However, the two antagonists together produced strong disinhibition that resembled the effect of removing the antennae (Fig. 5).

## Discussion

Many previous studies have shown that odours can inhibit spiking in olfactory bulb mitral cells and antennal lobe PNs (see refs 27–32 for early examples), but in principle this inhibition could be purely intraglomerular[15–17]. Experiments *in vitro* have revealed several types of interglomerular circuits[33–35], but some of these circuits are evidently not recruited by olfactory stimuli[15]. Here we have directly demonstrated an important role for inhibitory interactions between olfactory glomeruli *in vivo*.

Our results argue that a substantial component of interglomerular inhibition occurs at a presynaptic locus. Previous studies in other species have shown that GABA can inhibit release from ORN axon terminals[15–21]. Our results imply that in *Drosophila* this is mediated by both $GABA_A$ and $GABA_B$ receptors. This arrangement is unusual but not unique; for example, there are several instances of $GABA_A$ and $GABA_B$ inhibition at the same presynaptic site in other neural circuits[36–39]. Ionotropic and metabotropic receptors act with different kinetics, and so this arrangement might ensure that inhibition spans a broad time window. Although both receptor types were co-active during most of the odour response, we noticed that $GABA_A$ receptors were required for a brief early phase of inhibition after odour onset (Fig. 5) whereas $GABA_B$ receptors were required for the long, late phase (Figs 3 and 5).

We have shown that lateral inhibitory input to a glomerulus roughly scales with total ORN activity. This would imply that the odour tuning of GABAergic input to each glomerulus is approximately similar. However, the effects of lateral input may nevertheless be somewhat glomerulus-specific. Even if the odour tuning of GABA release were identical in all glomeruli, the efficacy of presynaptic inhibition will vary with presynaptic membrane potential[40,41]. As a result, the same inhibitory signal should be more effective in some glomeruli than in others. Also, in some glomeruli, lateral inhibition might be outweighed by lateral excitation. This would explain why other studies have found that some PNs can be excited by odours that do not excite their cognate ORNs[8,10,12,42].

We propose that this form of gain control represents a flexible balance between sensitivity and efficiency. When total ORN input is weak, lateral inhibition is minimal, and ORN–PN synapses are strong. When an odorant recruits vigorous ORN input to many glomeruli, GABAergic interneurons inhibit ORN neurotransmitter release. This should prevent a stimulus from saturating the dynamic range of many types of PNs simultaneously. Because this mechanism suppresses responses that are strong and redundant, it may tend to decrease cross-correlations between the output of different glomeruli, and thus promote a more efficient neural code for odours.

## METHODS SUMMARY

*In vivo* whole-cell patch-clamp recordings from PNs and extracellular recordings from ORNs were performed essentially as described previously[8–11,42]. Recordings in Fig. 3a–c were made from VM2 PNs in *Or43b[1]* mutant flies. Antagonists (50 μM CGP54626 and 5 μM picrotoxin, PCT) were added to the saline, which perfused the brain. Pertussis toxin was expressed under control of the *Or83b-GAL4* driver. All data are shown as mean values, averaged across experiments, ± s.e.m. (except for raw electrophysiological traces and rasters). The odour stimulus period (shown as a black bar in Figures) was 500 ms, except in Fig. 3b, c.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 22 September 2007; accepted 25 February 2008.**
**Published online 16 March 2008.**

1. Laissue, P. P. *et al.* Three-dimensional reconstruction of the antennal lobe in *Drosophila melanogaster. J. Comp. Neurol.* 405, 543–552 (1999).

2. Stocker, R. F., Lienhard, M. C., Borst, A. & Fischbach, K. F. Neuronal architecture of the antennal lobe in *Drosophila melanogaster*. *Cell Tissue Res.* **262,** 9–34 (1990).
3. Couto, A., Alenius, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr. Biol.* **15,** 1535–1547 (2005).
4. Fishilevich, E. & Vosshall, L. B. Genetic and functional subdivision of the *Drosophila* antennal lobe. *Curr. Biol.* **15,** 1548–1553 (2005).
5. Marin, E. C., Jefferis, G. S., Komiyama, T., Zhu, H. & Luo, L. Representation of the glomerular olfactory map in the *Drosophila* brain. *Cell* **109,** 243–255 (2002).
6. Wang, J. W., Wong, A. M., Flores, J., Vosshall, L. B. & Axel, R. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112,** 271–282 (2003).
7. Root, C. M., Semmelhack, J. L., Wong, A. M., Flores, J. & Wang, J. W. Propagation of olfactory information in *Drosophila*. *Proc. Natl Acad. Sci. USA* **104,** 11826–11831 (2007).
8. Wilson, R. I., Turner, G. C. & Laurent, G. Transformation of olfactory representations in the *Drosophila* antennal lobe. *Science* **303,** 366–370 (2004).
9. Wilson, R. I. & Laurent, G. Role of GABAergic inhibition in shaping odor-evoked spatiotemporal patterns in the *Drosophila* antennal lobe. *J. Neurosci.* **25,** 9069–9079 (2005).
10. Bhandawat, V., Olsen, S. R., Gouwens, N. W., Schlief, M. L. & Wilson, R. I. Sensory processing in the *Drosophila* antennal lobe increases reliability and separability of ensemble odor representations. *Nature Neurosci.* **10,** 1474–1482 (2007).
11. Olsen, S. R., Bhandawat, V. & Wilson, R. I. Excitatory interactions between olfactory processing channels in the *Drosophila* antennal lobe. *Neuron* **54,** 89–103 (2007).
12. Shang, Y., Claridge-Chang, A., Sjulson, L., Pypaert, M. & Miesenbock, G. Excitatory local circuits and their implications for olfactory processing in the fly antennal lobe. *Cell* **128,** 601–612 (2007).
13. Stocker, R. F., Heimbeck, G., Gendre, N. & de Belle, J. S. Neuroblast ablation in *Drosophila* P[GAL4] lines reveals origins of olfactory interneurons. *J. Neurobiol.* **32,** 443–456 (1997).
14. Ng, M. *et al.* Transmission of olfactory information between three populations of neurons in the antennal lobe of the fly. *Neuron* **36,** 463–474 (2002).
15. McGann, J. P. *et al.* Odorant representations are modulated by intra- but not interglomerular presynaptic inhibition of olfactory sensory neurons. *Neuron* **48,** 1039–1053 (2005).
16. Murphy, G. J., Darcy, D. P. & Isaacson, J. S. Intraglomerular inhibition: signaling mechanisms of an olfactory microcircuit. *Nature Neurosci.* **8,** 354–364 (2005).
17. Vucinic, D., Cohen, L. B. & Kosmidis, E. K. Interglomerular center-surround inhibition shapes odorant-evoked input to the mouse olfactory bulb *in vivo*. *J. Neurophysiol.* **95,** 1881–1887 (2006).
18. Nickell, W. T., Behbehani, M. M. & Shipley, M. T. Evidence for $GABA_B$-mediated inhibition of transmission from the olfactory nerve to mitral cells in the rat olfactory bulb. *Brain Res. Bull.* **35,** 119–123 (1994).
19. Wachowiak, M. & Cohen, L. B. Presynaptic inhibition of primary olfactory afferents mediated by different mechanisms in lobster and turtle. *J. Neurosci.* **19,** 8808–8817 (1999).
20. Aroniadou-Anderjaska, V., Zhou, F. M., Priest, C. A., Ennis, M. & Shipley, M. T. Tonic and synaptically evoked presynaptic inhibition of sensory input to the rat olfactory bulb via $GABA_B$ heteroreceptors. *J. Neurophysiol.* **84,** 1194–1203 (2000).
21. Wachowiak, M. *et al.* Inhibition of olfactory receptor neuron input to olfactory bulb glomeruli mediated by suppression of presynaptic calcium influx. *J. Neurophysiol.* **94,** 2700–2712 (2005).
22. Distler, P. G. & Boeckh, J. Synaptic connections between identified neuron types in the antennal lobe glomeruli of the cockroach, Periplaneta americana: II. Local multiglomerular interneurons. *J. Comp. Neurol.* **383,** 529–540 (1997).
23. Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125,** 143–160 (2006).
24. Zucker, R. S. & Regehr, W. G. Short-term synaptic plasticity. *Annu. Rev. Physiol.* **64,** 355–405 (2002).
25. Larsson, M. C. *et al. Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43,** 703–714 (2004).
26. Ferris, J., Ge, H., Liu, L. & Roman, G. G(o) signaling is required for *Drosophila* associative learning. *Nature Neurosci.* **9,** 1036–1040 (2006).
27. Shibuya, T., Ai, N. & Takagi, S. F. Response types of single cells in the olfactory bulb. *Proc. Jpn. Acad.* **38,** 231–233 (1962).
28. Macrides, F. & Chorover, S. L. Olfactory bulb units: activity correlated with inhalation cycles and odor quality. *Science* **175,** 84–87 (1972).
29. Mathews, D. F. Response patterns of single units in the olfactory bulb of the rat to odor. *Brain Res.* **47,** 389–400 (1972).
30. Tanabe, T., Iino, M. & Takagi, S. F. Discrimination of odors in olfactory bulb, pyriform-amygdaloid areas, and orbitofrontal cortex of the monkey. *J. Neurophysiol.* **38,** 1284–1296 (1975).
31. Meredith, M. & Moulton, D. G. Patterned response to odor in single neurones of goldfish olfactory bulb: influence of odor quality and other stimulus parameters. *J. Gen. Physiol.* **71,** 615–643 (1978).
32. Chaput, M. & Holley, A. Single unit responses of olfactory bulb neurones to odour presentation in awake rabbits. *J. Physiol. (Paris)* **76,** 551–558 (1980).
33. Isaacson, J. S. & Strowbridge, B. W. Olfactory reciprocal synapses: dendritic signaling in the CNS. *Neuron* **20,** 749–761 (1998).
34. Urban, N. N. & Sakmann, B. Reciprocal intraglomerular excitation and intra- and interglomerular lateral inhibition between mouse olfactory bulb mitral cells. *J. Physiol. (Lond.)* **542,** 355–367 (2002).
35. Aungst, J. L. *et al.* Centre-surround inhibition among olfactory bulb glomeruli. *Nature* **426,** 623–629 (2003).
36. Stuart, G. J. & Redman, S. J. The role of $GABA_A$ and $GABA_B$ receptors in presynaptic inhibition of Ia EPSPs in cat spinal motoneurones. *J. Physiol. (Lond.)* **447,** 675–692 (1992).
37. Matthews, G., Ayoub, G. S. & Heidelberger, R. Presynaptic inhibition by GABA is mediated via two distinct GABA receptors with novel pharmacology. *J. Neurosci.* **14,** 1079–1090 (1994).
38. Fischer, Y. & Parnas, I. Differential activation of two distinct mechanisms for presynaptic inhibition by a single inhibitory axon. *J. Neurophysiol.* **76,** 3807–3816 (1996).
39. Fischer, Y. & Parnas, I. Activation of $GABA_B$ receptors at individual release boutons of the crayfish opener neuromuscular junction produces presynaptic inhibition. *J. Neurophysiol.* **75,** 1377–1385 (1996).
40. Bean, B. P. Neurotransmitter inhibition of neuronal calcium currents by changes in channel voltage dependence. *Nature* **340,** 153–156 (1989).
41. Foldy, C., Neu, A., Jones, M. V. & Soltesz, I. Presynaptic, activity-dependent modulation of cannabinoid type 1 receptor-mediated inhibition of GABA release. *J. Neurosci.* **26,** 1465–1469 (2006).
42. Schlief, M. L. & Wilson, R. I. Olfactory processing and behavior downstream from highly selective receptor neurons. *Nature Neurosci.* **10,** 623–630 (2007).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank K. Ito, L. Luo, G. Roman, D. P. Smith, L. M. Stevens and L. B. Vosshall for gifts of fly stocks. We thank G. Laurent, A. W. Liu and members of the Wilson laboratory for conversations. This work was funded by a grant from the NIDCD, the Pew, McKnight, Sloan, and Beckman Foundations (to R.I.W.). S.R.O. was partially supported by a NSF fellowship.

**Author Contributions** S.R.O. performed the experiments and analysed the data. S.R.O. and R.I.W. designed the experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.I.W. (rachel_wilson@hms.harvard.edu).

## METHODS

**Fly stocks.** Flies were reared at 20–21 °C on conventional cornmeal agar. All experiments were performed on adult female flies 2–5 days post-eclosion. Fly stocks were provided as follows: *Or43b[1]* (D. Smith); *NP5103-GAL4*, *NP3481-GAL4* and *NP5221-GAL4* (K. Ito and L. Luo); *Or46a-GAL4* and *Or83b-GAL4* (L. Vosshall); $UAS\text{-}DTl_{II}$ and $UAS\text{-}DTl_{III}$ (L. Stevens); *UAS-PTX* (G. Roman); and $UAS\text{-}CD8GFP_I$ (Bloomington Stock Center).

**ORN recordings.** Extracellular recordings of ORN spiking were performed as described previously[10]. To accurately sort spikes recorded from VC1 ORNs, we killed the second ORN type (VA7l) housed in the same sensillum. This was done by expressing diphtheria toxin light chain in the VA7l ORNs (genotype: either $Or46a\text{-}GAL4/UAS\text{-}DTl_{II}$ ($n = 4$) or $Or46a\text{-}GAL4/+; UAS\text{-}DTl_{III}/+$ ($n = 4$)). The genotype for VM7 ORN recordings was *NP3481-Gal4 UAS-CD8GFP*, which is the same genotype as that used for the VM7 PN recordings. VM7 ORNs fire spontaneously at 10 spikes $s^{-1}$, and VC1 ORNs fire spontaneously at 3 spikes $s^{-1}$.

**Projection neuron recordings.** *In vivo* whole-cell recordings from PNs were performed as previously described[10]. For the current-clamp recordings in Figs 1, 2 and 5 and in Supplementary Figs 2–5, the composition of the internal patch-pipette solution was (in mM): potassium aspartate 140, HEPES 10, MgATP 4, $Na_3GTP$ 0.5, EGTA 1, KCl 1, and biocytin hydrazide 13 (pH 7.3, osmolarity adjusted to ~265 mOsm). For the voltage-clamp recordings in Figs 3 and 4 and Supplementary Figs 6–8, the composition of the internal was (in mM): cesium aspartate 140, HEPES 10, MgATP 4, $Na_3GTP$ 0.5, EGTA 1, KCl 1, biocytin hydrazide 13, and QX-314 10 (pH 7.3, osmolarity adjusted to ~265 mOsm). The composition of the external saline solution in all recordings was (in mM): NaCl 103, KCl 3, *N*-tris(hydroxymethyl) methyl-2-aminoethane-sulfonic acid 5, trehalose 8, glucose 10, $NaHCO_3$ 26, $NaH_2PO_4$ 1, $CaCl_2$ 1.5, and $MgCl_2$ 4. Osmolarity was adjusted to 270–275 mOsm. The saline was bubbled with 95% $O_2$/5% $CO_2$ and reached a final pH of 7.3. Saline perfused the brain continuously at 2 ml $min^{-1}$. Recordings were obtained with an A-M Systems Model 2400 amplifier (100 MΩ headstage), were low-pass filtered at 5 kHz, and digitized at 10 kHz. Data were acquired in Igor Pro. In most experiments, we used an enhancer trap line to label specific PNs with green fluorescent protein (GFP) for targeted recording. VM7 PNs were labelled by *NP3481-Gal4 UAS-CD8GFP* (Figs 1, 2 and 5, and Supplementary Figs 2, 4 and 5). VC1 PNs were labelled by *NP5221-Gal4 UAS-CD8GFP* (Fig. 1 and Supplementary Fig. 3). VM2 PNs were labelled by *NP5103-Gal4 UAS-CD8GFP* (Fig. 3a–c, and Supplementary Fig. 6a–c and the 'odour' condition in Supplementary Fig. 7d). VM2 PNs were recorded in flies in which the odorant receptor expressed by VM2 ORNs was mutated (*NP5103-Gal4 UAS-CD8GFP; Or43b[1]*, see schematic in Fig. 3a). In some experiments for each genotype, we filled the recorded cell with biocytin and verified the PN identity with biocytin histochemistry as described previously[8]. In the GABA iontophoresis experiments in Fig. 3d–f (also Supplementary Figs 6d–f and 7), we recorded from PNs in the anterodorsal cell cluster[5] labelled by the enhancer trap line *GH146* (*GH146-Gal4 UAS-CD8GFP*). For the experiments in which pertussis toxin was selectively expressed in ORNs (Fig. 4, also Supplementary Fig. 8), we recorded from PNs in the anterodorsal cell cluster in the genotype *Or83b-Gal4; UAS-PTX/+*. In voltage-clamp recordings, the command potential was either −85 mV (Fig. 3a–c) or −65 mV (Figs 3d–f and 4).

**Manipulation of peripheral organs.** Antennal input was abolished in some experiments by severing the antennal nerves with fine forceps just before recording. The antennal nerve was gently broken by applying forces perpendicular to the long axis of the nerve, leaving a stump of nerve attached to the antennal lobe. In other experiments, the maxillary palps were removed with forceps just before recording. For both antennal and palp amputation, the cell bodies of the affected ORNs are removed, but the proximal portions of their axons remain intact and continue to innervate the antennal lobe. Amputation removes ORN somata and thus large spontaneous spike-driven EPSCs (~5–20 pA) disappear almost entirely. Axon terminals are still intact, as evidenced by the persistence of miniature EPSCs (~1 pA). In some experiments, the maxillary palps were shielded from direct odour stimulation by covering them in '5 minute epoxy' (Devcon). The epoxy was allowed to dry for ~20 min before the recording was started. We confirmed that this substance is relatively non-toxic to ORNs by covering the antennae in epoxy, allowing it to dry, peeling the epoxy away, and verifying that odours were still able to elicit a normal field-potential response in the antennae.

**Olfactory stimulation.** Odours were diluted in paraffin oil at a ratio of 1:100 v/v. Odour dilutions in paraffin oil were prepared fresh every 5 days. Our odour panel consisted of benzaldehyde, butyric acid, 1-butanol, cyclohexanone, ethyl butyrate, ethyl acetate, ethyl-3-hydroxybutyrate, fenchone, geranyl acetate, 2-heptanone, linalool, 4-methyl cyclohexanol, methyl salicylate, 3-methylthio-1-propanol, 4-methyl phenol, octanal, 1-octen-3-ol, pentyl acetate, trans-2-hexenal and paraffin oil (solvent control). Odour source details are posted at http://wilson.med.harvard.edu/odors.html. Odours were delivered with a custom-built olfactometer as described previously[10]. Odour stimuli were applied for 500 ms, except in Fig. 3 in which the odour was applied for 1 s. For the experiments in Figs 1, 2 and 5, consecutive odour presentations were spaced 25–35 s apart.

**Direct ORN axon stimulation.** In Figs 3 and 4 and in Supplementary Figs 6–8, we electrically stimulated ORN axons. The ipsilateral antennal nerve was severed and inserted into a stimulating suction electrode. Single spikes were evoked in ORN axons with a short current pulse (50 μs) using a stimulus isolator (A.M.P.I.). In each sweep we delivered a train of 23 nerve stimuli at 4 Hz. During the train, the size of the evoked EPSCs decreased over the first 4–8 stimuli owing to short-term depression, but then reached a steady value. We timed our GABA iontophoresis and odour stimulation to occur during this steady-state period. For the paired-pulse experiments in Supplementary Fig. 7, we delivered two electrical stimuli to the nerve with an inter-pulse interval of 25 ms. This interval produces depression at the ORN–PN synapse under control conditions and allowed us to observe increases in the paired-pulse ratio when we decreased release probability.

**GABA iontophoresis.** In the GABA iontophoresis experiments, a high-resistance (~80 ΩM) sharp glass pipette was filled with a solution of 250 mM GABA in water, and the pH was adjusted to 4.3 by adding HCl. GABA was ejected into the antennal lobe neuropil with a brief (3–20 ms) positive current pulse using an iontophoresis unit (World Precision Instruments). A constant negative backing current was applied to retain GABA in the iontophoresis pipette between ejection events. To ensure that a similar amount of GABA was released into the neuropil in all experiments, and that the time-course of GABA release was as consistent as possible, we adjusted the location of the iontophoresis pipette, the level of negative backing current and the duration of the ejection current at the beginning of each experiment. We did not proceed with an experiment until the following criteria were met: EPSC suppression was ~50% at 150 ms after iontophoresis, maximal EPSC suppression was ~85% at 400 ms after iontophoresis, and GABA evoked an outward current in the recorded PN which lasted for ~1–1.5 s.

**Antagonists.** Antagonists were prepared as concentrated stock solutions and a measured volume of stock was added to the saline perfusate to achieve the final drug concentrations. Picrotoxin (Sigma) was used at 5 μM, CGP54626 (Tocris) was used at 50 μM, and $CdCl_2$ (Sigma) was used at 10–25 μM.

**Peri-stimulus time histograms.** For the experiments in Figs 1 and 5 and in Supplementary Figs 2 and 3, spike times were extracted from raw ORN and PN recordings using routines in Igor Pro. Each cell was tested with multiple odours, and each odour was presented six times. The response to the first presentation was not included in our analysis. Each of the five remaining trials was converted into a peri-stimulus time histogram (PSTH) by counting the number of spikes in 50-ms bins that overlapped by 25 ms. These single-trial PSTHs were averaged together to generate a PSTH describing the response to an odour in a given experiment. Multiple cells corresponding to each glomerular class and each cell type (ORN or PN) were tested with a given odour in multiple experiments. Average PSTHs in Fig. 1 and Supplementary Figs 2 and 3 represent the mean ± s.e.m. computed across experiments.

**Input–output functions.** For the analysis of odour responses in Fig. 1d, we computed the average spike rate over the 500-ms odour duration and subtracted the baseline firing rate averaged over the 500-ms window before odour onset. Each point in the input–output function represents the mean PN response to an odour plotted against the mean response of the cognate ORNs to the same odour. Fits are single exponential functions.

**Lifetime sparseness.** The selectivity of a neuron's odour response profile (Fig. 1e) was quantified as lifetime sparseness:

$$S = \frac{1}{1 - 1/N}\left(1 - \frac{\left(\sum_{j=1}^{N} r_j/N\right)^2}{\sum_{j=1}^{N} r_j^2/N}\right)$$

where $N$ = the number of odours and $r_j$ is the analogue response intensity of the neuron to odour $j$ minus baseline firing rate. Analogue response intensity was the mean spike rate (averaged across five sweeps) during the entire 500-ms odour stimulus period. Any values of $r_j < 0$ were set to zero before computing lifetime sparseness (this was the only analysis in this study in which negative responses were zeroed). Responses to paraffin oil (solvent control) were not considered in the sparseness analysis. We computed the lifetime sparseness for each individual cell in our study for which we tested at least 12 of the odours in our set. We used the Mann–Whitney $U$-test to assess the significance of sparseness differences.

**Correlations between ranked odour preferences.** To compute correlations between ORN and PN odour preferences, we computed Spearman's rank correlation coefficient ($\rho$) on the odour response vectors of these neurons. This is a

nonparametric measure that quantifies how well any monotonic function could describe the relationship between ORN and PN responses. We chose this rather than a linear (Pearson's) correlation coefficient because the intrinsic input–output function of these glomeruli is highly nonlinear (Fig. 1d). For each glomerulus, we computed Spearman's $\rho$ on each pairwise combination of individually recorded ORNs and PNs for which we tested at least 12 of the same odours on both cells (for glomerulus VM7, $n = 27$ comparisons for ORNs compared to PNs with antennae and $n = 38$ for ORNs compared to PNs without antennae; for glomerulus VC1, $n = 19$ for ORNs compared to PNs with antennae and $n = 25$ for ORNs compared to PNs without antennae). We report the mean of these pairwise correlations. We assessed the significance of differences in Spearman's $\rho$ with the Mann–Whitney *U*-test.

**Quantifying changes in membrane potential.** In Figs 2c and 5c, we show membrane potential responses averaged across multiple experiments. To generate these we first averaged 4–6 consecutive sweeps within a single experiment for each odour and low-pass filtered this at 13 Hz to remove any spikes. Next we averaged these responses across experiments, with ± s.e.m. shown in pastel. To quantify the magnitude of lateral input in Fig. 2d–f, we integrated the change in membrane potential (versus baseline) over a 2-s interval beginning at odour onset. We chose this interval because shorter intervals did not capture as much odour dependence in the magnitude of lateral input. In Fig. 2d, f, PN disinhibition was computed by averaging PN firing rates over the 2 s after odour onset, and then averaging across experiments (4–12 PNs per odour), before computing the difference in the PN responses with antennae versus without antennae. In Fig. 2e, total ORN activity for each odour was computed by summing all the response rates in ref. 23. There are only 14 data points in this panel because only 14 of our odours were included in ref. 23.

**EPSC amplitudes.** We measured EPSC amplitudes in Figs 3 and 4 and in Supplementary Figs 6–8 by first averaging together a block of 8–20 sweeps from the same experiment. From these averaged traces, we computed the EPSC amplitude as the average over a 0.9-ms window centred on the peak. Before averaging across experiments, we normalized the train of evoked EPSCs to the baseline EPSCs evoked just before either odour or GABA onset.

ARTICLES

# Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA

Cai-Guang Yang[1]*, Chengqi Yi[1]*, Erica M. Duguid[1], Christopher T. Sullivan[1], Xing Jian[1], Phoebe A. Rice[2] & Chuan He[1]

**_Escherichia coli_ AlkB and its human homologues ABH2 and ABH3 repair DNA/RNA base lesions by using a direct oxidative dealkylation mechanism. ABH2 has the primary role of guarding mammalian genomes against 1-meA damage by repairing this lesion in double-stranded DNA (dsDNA), whereas AlkB and ABH3 preferentially repair single-stranded DNA (ssDNA) lesions and can repair damaged bases in RNA. Here we show the first crystal structures of AlkB–dsDNA and ABH2–dsDNA complexes, stabilized by a chemical cross-linking strategy. This study reveals that AlkB uses an unprecedented base-flipping mechanism to access the damaged base: it squeezes together the two bases flanking the flipped-out one to maintain the base stack, explaining the preference of AlkB for repairing ssDNA lesions over dsDNA ones. In addition, the first crystal structure of ABH2, presented here, provides a structural basis for designing inhibitors of this human DNA repair protein.**

Cellular DNA is constantly subjected to modifications by environmental and endogenous chemicals, which can result in covalent changes[1,2]. Methylating (or alkylating) agents are a common group of DNA modifiers that introduce damage primarily to the heterocyclic bases of DNA, with mutagenic and/or cytotoxic consequences. Alkylating agents are also widely used in cancer therapy and exert anticancer effects by creating cytotoxic DNA lesions in tumour cells. Many of these alkylation DNA damages are detected and repaired by proteins that are conserved across kingdoms.

The *E. coli* AlkB protein is a direct dealkylation DNA repair protein[3–5]. It uses a mononuclear iron(II) site and cofactors 2-ketoglutarate (2KG) and dioxygen to perform an unprecedented oxidative demethylation of DNA base lesions 1-meA, 3-meC, 1-meG and 3-meT (Supplementary Fig. 2)[6–11]. AlkB also removes etheno DNA lesions by using a similar oxidation mechanism[12,13]. There are nine potential human homologues of AlkB. Two of these, ABH2 and ABH3, can repair the same spectrum of base lesions as AlkB[14–16]. A third one, FTO, has been recently shown to repair 3-meT in DNA, and it affects obesity in mammals through an unknown mechanism[17]. ABH2 acts as the primary housekeeping enzyme in mammals for repairing endogenously formed 1-meA lesions in duplex DNA[18]. The *in vivo* function of ABH3 is still unclear but it has been shown to repair base lesions in RNA as well as DNA[15]. Although the structures of AlkB (with dT-(1-medA)-dT)[19] and ABH3 (in the absence of DNA or RNA)[20] were solved recently, the molecular mechanism underlying the different substrate preferences of ABH2 and AlkB remains unknown. Furthermore, the structure of ABH2 has, to our knowledge, never been reported.

## Cross-linking to stabilize protein–DNA complexes

We report here the first crystal structures of AlkB–dsDNA and ABH2–dsDNA complexes. The AlkB family proteins bind DNA weakly[21] and form labile complexes with damage-containing DNA[22], which makes crystallization of their protein–DNA complexes challenging. To overcome this difficulty we used chemical cross-linking methods[23,24]; initially using an active site disulphide cross-linking strategy that we developed previously (Fig. 1a)[25,26]. Base-repair proteins flip damaged bases and insert them into the active site for processing. Therefore, we reasoned, a cysteine residue engineered into the active site of AlkB may form a disulphide cross-link, at equilibrium, with a disulphide-modified cytosine (C* in a C*:A base pair) flipped into the active site of the repair protein (Fig. 1a)[27]. Of several mutations tested, AlkB D135C and ABH2 E175C gave the best cross-linking yields with disulphide-modified DNA under equilibrium conditions (Supplementary Fig. 3).

## Structure determination

For crystallization, truncated but catalytically active AlkB-ΔN11[19] and ABH2-ΔN55 were used (Supplementary Fig. 4). Of multiple DNA sequences screened, a 13-base-pair duplex DNA1, with a central C*:A base pair (Fig. 1a), gave the best crystals of the active site cross-linked AlkB–DNA complex. These crystals diffracted to 1.65 Å and the structure was solved by molecular replacement using the previously reported AlkB structure as the search model[19]. Two crystal forms of active site cross-linked ABH2–DNA complexes were grown using different 13-base-pair dsDNAs containing central C*:A pairs. The first, with DNA2, diffracted to 2.0 Å, and was solved by multi-wavelength anomalous dispersion (MAD) phasing using bromine-labelled (5-Br dU) DNA. The second form, with DNA3, diffracted to 3.0 Å and was solved by molecular replacement using the first as a search model. Both the AlkB and the ABH2 truncations used in this study still efficiently repair 1-meA in dsDNA with the sequences shown in the crystal structures (Supplementary Figs 5 and 6).

## General features of the protein–DNA interactions

AlkB and ABH2 have similar folds (a common jelly-roll fold that is shared by all iron(II)/2-KG-dependent oxygenases[28–30]), and both flip the damaged base into a deep active site pocket. However, their interactions with dsDNA are very different (Fig. 1b–e). First, AlkB

[1]Department of Chemistry, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. [2]Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.
*These authors contributed equally to this work.

interacts almost exclusively with the damaged strand, whereas ABH2 makes significant contacts with both DNA strands (Supplementary Fig. 7). Second, the two enzymes use very different mechanisms to flip the damaged base. ABH2 uses an aromatic finger residue, Phe 102, which resembles the intercalating hairpin motif observed with DNA glycosylases[31,32], to intercalate into the duplex stack and fill the DNA gap resulting from the base flipping. AlkB lacks this finger residue; instead it manipulates the DNA backbone near the flipped base such that the bases flanking it stack with one another. Two slightly different DNA conformations were observed in the AlkB–DNA1 structure, but in both cases a similar distortion of the lesion-containing strand was observed (Fig. 1b, d, and Supplementary Fig. 8).

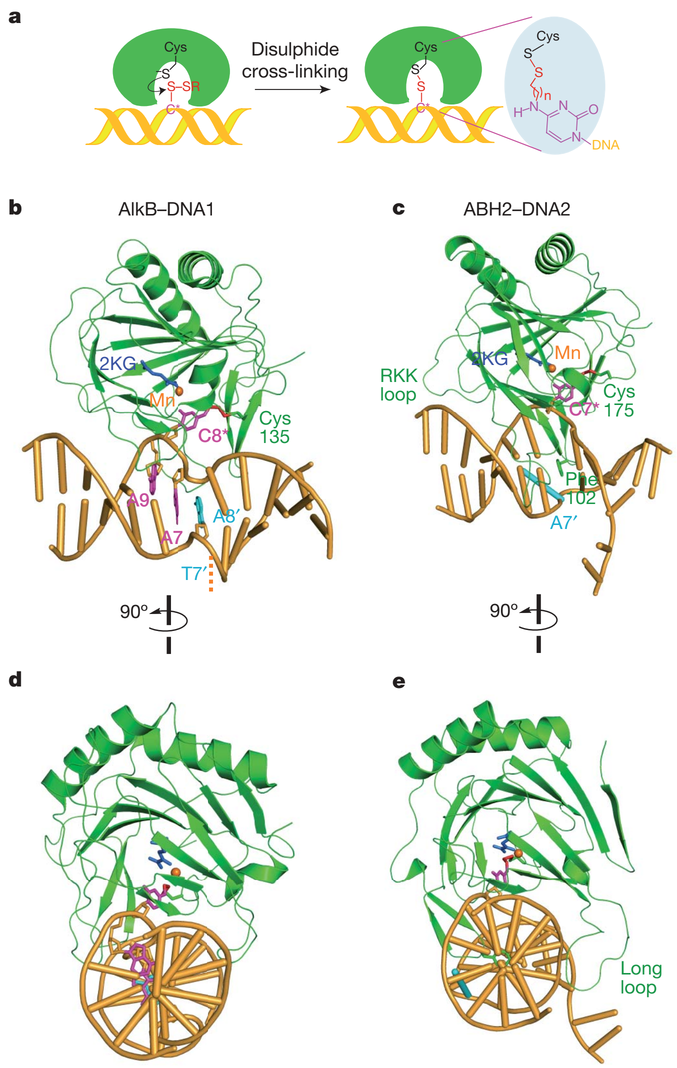



**Figure 1 | Crystal structures of the active site cross-linked AlkB–DNA1 and ABH2–DNA2 complexes. a**, The active site cross-linking strategy. A disulphide-modified cytosine C* (*n* = 1 or 2) in a mismatched C*:A base pair can be flipped into the active site of AlkB or ABH2. A covalent disulphide bond can form between C* and an engineered Cys in the active site of the repair protein under equilibrium conditions. **b**, Cartoon of the AlkB–DNA1 complex. Two slightly different DNA conformations are observed in this structure; one is shown here, the other in Supplementary Fig. 8. The protein is shown in green (the cross-linking residue Cys 135 is labelled), manganese(II) in orange, 2KG in blue, DNA in yellow–orange, the flipped base C8* and the two bases flanking C8* in light magenta, the disulphide bond in red and the orphaned base A8′ in cyan. The approximate location of the disordered base T7′ is shown as a dotted line. **c**, Cartoon trace of the ABH2–DNA2 complex. The same colour coding as in **b** is used. The finger residue Phe 102 and the cross-linking residue Cys 175 are green, and the orphaned base A7′ is cyan. The location of a DNA-binding loop containing a RKK sequence is indicated. **d**, **e**, Same structure as in **b** and **c**, rotated 90° to the right. A DNA-binding flexible long loop is labelled. See Supplementary Fig. 15 for a topological diagram of ABH2.

## Cross-linking AlkB to dsDNA away from the active site

The protein portion of the AlkB–DNA1 structure overlaps almost perfectly with the previously published AlkB structure (Supplementary Fig. 9a). However, to confirm that the DNA distortions observed are not due to cross-linking to the active site of the protein, we also engineered a disulphide cross-link away from the active site, to AlkB S129C (Fig. 2a)[32,33]. Two such AlkB–dsDNA complexes were crystallized: one with DNA4, a 13-base-pair dsDNA containing C10*, 1-meA and an abasic site (Fig. 2b); and one with DNA5, containing C10* two base pairs away from the 1-meA (Fig. 2c). These were solved at 1.68 Å and 1.90 Å, respectively, by molecular replacement.

## Unique binding of duplex DNA by AlkB

All three structures of the AlkB–dsDNA complexes show a base-flipping feature distinct from all known structures of base-flipping proteins bound to DNA. The two bases that flank the flipped 1-meA, A7 and A9, are 'squeezed' such that they stack on one another, about 3.4 Å apart (Fig. 3a and Supplementary Fig. 10c). The sugar rings of these two nucleosides adopt unusual conformations. One (A9) is C3′-endo, standard in A-form DNA (Supplementary Fig. 10a). The other (A7) shows the most surprising feature: with its base, it is twisted by about 180°, resulting in an inverted sugar (O4′ points towards the 3′ end of the DNA strand) while the base stacks with A9:T9′ (Fig. 3a). The inversion of this sugar ring is locally enforced by residues 51–53 (Thr-Pro-Gly), which form a rigid hydrogen bonding network with the phosphate group between A7 and A6 (Fig. 3a and Supplementary Fig. 10b), and through multiple protein–DNA interactions anchoring the two phosphate groups flanking the 1-meA (Supplementary Fig. 7a). These features were not observed in the AlkB structure containing a short T-(1-meA)-T[19], most likely

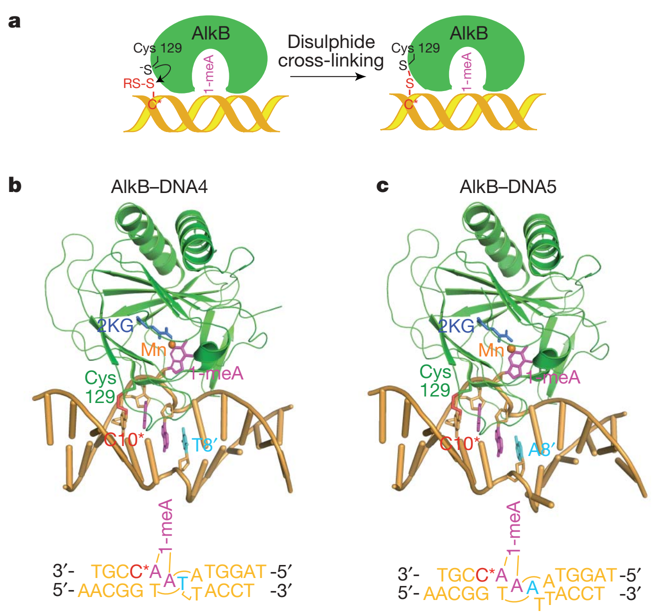



**Figure 2 | Crystal structures of AlkB–DNA4 and AlkB–DNA5 complexes with 1-meA recognized by an intact active site. a**, A strategy to set up a disulphide cross-link away from the active site of AlkB. **b**, Cartoon of the AlkB–DNA4 complex with 1-meA (light magenta) recognized in the active site of AlkB. The complex was stabilized by a disulphide link between an engineered Cys 129 (green) and C10* that is two base pairs away from 1-meA. The colour coding is the same as in Fig. 1b. DNA4 has an abasic site '␣' (position 7′) 3′ to T8′ of the 1-meA:T base pair. **c**, Cartoon trace of the AlkB–DNA5 complex; DNA5 has a central 1-meA:A base pair replacing a C*:A in DNA1 and a C10* cross-linking site two base pairs away from 1-meA.

because the ssDNA trimer is too short to reach much of the DNA-binding surface of AlkB. As a result, the positions of the two bases that flank the 1-meA are different in the previous AlkB structure and the current AlkB–dsDNA structures (Supplementary Fig. 9c). It remains to be seen how a long ssDNA interacts with AlkB.

In all three AlkB–dsDNA structures, the protein stabilizes only the distorted conformation of the DNA immediately flanking the 1-meA; the complementary strand appears to accommodate this through spontaneous conformational rearrangements. The base opposite the 1-meA (T8′ in DNA4 or A8′ in DNA1 and DNA5) intercalates between A7 and A6 to generate a one-base-wide stack containing bases from both strands, which maintains the length of the normal duplex DNA (Fig. 3a and Supplementary Fig. 10c). The base at the 7′ position is forced out of the helix and is disordered in our structures (in DNA4 it is replaced with an abasic site) (Figs 1b and 2b, c). To support the observation that the complementary strand distorts to accommodate base flipping in AlkB–dsDNA complexes, we measured the context dependence of AlkB-mediated 1-meA repair (Supplementary Fig. 11). AlkB can catalytically repair this lesion in the dsDNA sequence used in our crystals (replacing C8*:A8′ with 1-meA:T in DNA1), which has two A:T base pairs flanking 1-meA:T. When the two flanking base pairs (7:7′ and 9:9′) were replaced with G:C, repair activity dropped approximately threefold. The activity of this substrate was restored to the original level by incorporating an abasic site at position 7′, which destabilizes the base pair containing the base that is forced out of the double helix in our structures. The highest catalytic repair activity was observed with the ssDNA strand containing 1-meA, as expected.

## ABH2 binding to duplex DNA

The structure of ABH2 differs from AlkB in having extra DNA-binding motifs that grasp the complementary strand of the duplex DNA (Fig. 1). ABH2 uses a short loop with a positively charged RKK sequence (Arg 241–Lys 243) and an additional long, flexible loop carrying DNA-binding residues Arg 198, Gly 204 and Lys 205 to bind the opposite DNA strand (Fig. 1c, e and Supplementary Fig. 7b). The two bases that flank the flipped out C7*, A8 and T6, are approximately the same distance from each other as they would be in canonical duplex DNA (Fig. 3b). The orphaned base A7′ is flexible and poorly ordered. Because the normal duplex DNA length is maintained by the intercalating finger Phe 102, no further distortion of the DNA duplex is induced by ABH2.

To capture the structure of ABH2 with an intact active site, we engineered a disulphide cross-link away from the active site, to ABH2 G169C[32,33]. The cross-linked complex with a new 13-base-pair duplex probe containing C9* (DNA6; Fig. 4a and Supplementary Fig. 12) was crystallized and the structure was solved at 2.35 Å by molecular replacement. A cofactor-containing structure was also determined (at 2.6 Å) by soaking these crystals in 2KG and $Mn^{2+}$ (which occupies $Fe^{2+}$-binding site but does not support catalysis). The cofactors are bound by ABH2 in a standard conformation resembling that seen in AlkB and other iron(II)/2KG-based oxygenases[19,28]. The metal ion is ligated by His 171, Asp 173 and His 236, a bidentate 2KG and a water molecule in an octahedral geometry (Fig. 4b, and Supplementary Figs 13 and 14).

## Recognition of 1-meA by ABH2

The ABH2–dsDNA6 structure provides an undistorted view of the active site (Fig. 4b). A total of five active site residues are engaged in the recognition of 1-meA. Phe 124 and His 171 stack against the 1-meA; an analogous stacking interaction is also observed for 1-meA recognition by Trp 69 and His 131 in AlkB (Supplementary Fig. 9c). Three residues, Tyr 122, Glu 175 and Asp 174, together with a water molecule, form an extensive hydrogen-bonding network that contacts the N6 and N7 nitrogen atoms of 1-meA in ABH2. This tight recognition of 1-meA is not observed in the AlkB structure, as only two residues, Asp 135 and Glu 136 (mediated through a water molecule), were involved in the recognition of the N6 nitrogen of 1-meA by AlkB[19]. The positively charged 1-meA may be favourably recognized through electrostatic interaction with the nearby carboxylate of Glu 175 (Asp 135 of AlkB) and stacking with the π system of Phe 124 (Trp 69 of AlkB). These features may help ABH2 and AlkB discriminate against the neutral base A.

## Discussion

The AlkB family oxygenases represent a novel class of DNA base-repair enzymes that use an oxidative dealkylation mechanism. Certain members of this family, AlkB and ABH3, exhibit a distinct preference to repair ssDNA and RNA. We present the first structures

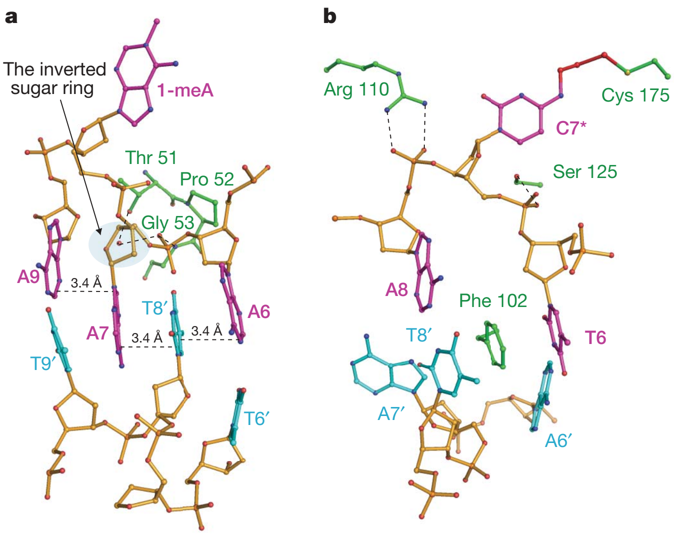



**Figure 3 | Close views of the base-flipping regions.** **a**, AlkB–DNA4 complex, with the same colouring as in Fig. 1b. A7 stacks with the A9:T9′ base pair. Thr 51 and Gly 53 are hydrogen bonded to the phosphate linking A7 and A6 (Supplementary Fig. 10b). The sugar ring of A9 adopts the 3′-endo conformation whereas the sugar ring of A7 is forced to be inverted by about 180° by the protein. T8′ from the complementary strand intercalates between A7 and A6, forming a one-base-wide stack containing bases from both strands. **b**, ABH2–DNA2 complex, with the same colour coding as in Fig. 1b. Phe 102 inserts into the DNA duplex with A7′ as the orphaned base.

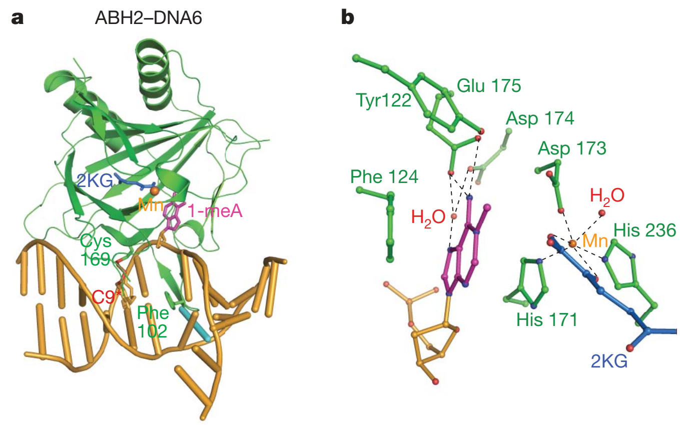



**Figure 4 | Crystal structure of the ABH2–DNA6 complex with 1-meA recognized by an intact active site.** **a**, Cartoon trace of the ABH2–DNA6 complex with 1-meA recognized in the active site of ABH2 and a disulphide cross-link (Cys 169 to C9*) installed two base pairs away from the flipped base. The same colour scheme is used as in Fig. 1b. **b**, View of the active site of ABH2 with manganese(II) (orange), 2KG (blue) and 1-meA (light magenta). N6 of 1-meA is within hydrogen-bond distance to Tyr 122 (3.3 Å) and Glu 175 (2.8 Å). N7 forms a hydrogen bond to a water (3.0 Å) which also interacts with Asp 174 (2.6 Å) and Glu 175 (2.6 Å).

of AlkB–dsDNA and ABH2–dsDNA complexes stabilized by chemical cross-linking, showing how AlkB and ABH2 bind and repair damage in dsDNA. The AlkB protein, lacking a finger residue, facilitates base flipping by squeezing together the two bases flanking the flipped-out one such that they stack on one another. This AlkB-mediated DNA distortion, together with the fact that AlkB mostly binds to the lesion-containing strand of DNA, explains the preference of AlkB for ssDNA substrates. The presence of the complementary strand in dsDNA rigidifies the duplex structure and serves as a 'non-competitive' inhibitor. The binding energy of complex formation must be used to pay an energetic penalty to flip 1-meA from a duplex DNA. A much lower energetic penalty would be associated with AlkB binding to the same base lesion in flexible ssDNA.

The ABH2–DNA structures clearly show ABH2 as a dsDNA repair protein, as the protein interacts extensively with both strands of dsDNA and uses a finger residue to intercalate in the DNA duplex to maintain continuous stacking. It has been commonly observed that single-base DNA glycosylases tend to contact only the lesion-containing strand of duplex DNA[31,34,35]. The extra capacity to interact with the other strand of DNA allows ABH2 to repair dsDNA lesions preferentially, which correlates well with its primary role as an enzyme that repairs 1-meA damage in mammalian genomes. Our structures also reveal a more complex arrangement of protein-(1-meA) interactions in the active site of ABH2 than AlkB, which provides a structural basis for designing small molecules that may shut down the ABH2-based repair activity potentially to improve efficacy of alkylation anticancer treatments.

Now with structures of AlkB–dsDNA, ABH2–dsDNA and ABH3 available, comparisons can be made (Fig. 5 and Supplementary Fig. 15). Surprisingly, ABH3, which shows substrate preferences similar to AlkB, shares some of the DNA-binding motifs of ABH2 in an overlay of these three structures. The long loop that binds the second DNA strand in ABH2 (between β11 and β12) is present in the sequence though disordered in the structure of ABH3 (between β10 and β11), but is much shorter in AlkB. Both ABH2 and ABH3 also contain a β-sheet hairpin motif (β3–β4 and β4–β5, respectively) that is absent in AlkB. In ABH2 this hairpin contains the aromatic finger residue, Phe 102, which is not conserved in ABH3. ABH3 also lacks the RKK motif used by ABH2 to contact the complementary strand of double-stranded DNA. How ABH3 may use these motifs to bind ssDNA or RNA preferentially remains to be seen in future structures of ABH3–DNA complexes.

Both AlkB and ABH2 are weak DNA-binding proteins. They likely interact with other DNA-binding proteins to stay close to DNA. ABH2, with a finger residue, may search for DNA damage by similar mechanisms proposed for DNA repair glycosylases[33,36]. AlkB, using a very different base-flipping mechanism, may impose a strain to the backbone of duplex DNA to facilitate identification of weakened base pairs, as suggested by the structures presented here. However, how does it locate damaged bases in flexible ssDNA? Further studies are required to answer these intriguing questions.

Lastly, the active site cross-linking strategy to trap labile protein–DNA complexes could be powerful for structurally characterizing other base-flipping proteins such as FTO and ABH3. As demonstrated here, once detailed protein–DNA interactions are revealed, a cross-link away from the active site can be readily engineered to provide native-like structures of the protein–DNA complexes.

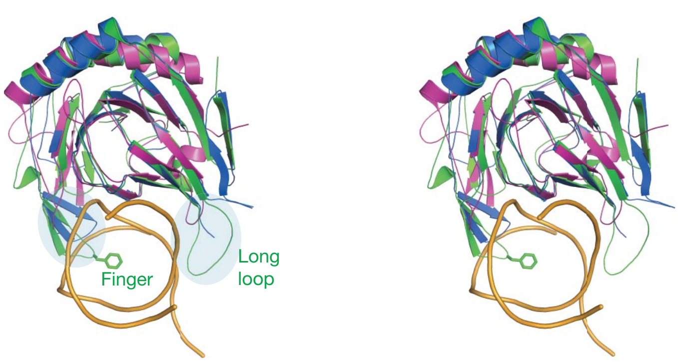



**Figure 5 | Structural comparison of AlkB, ABH2 and ABH3 (stereo view).** Least-squares superposition of structures of AlkB (in magenta, from AlkB–DNA1 without showing DNA1), ABH2–DNA2 (protein in green and DNA backbone in orange) and ABH3 (in blue, from Protein Data Bank 2IUW). The hairpin and flexible long-loop motifs are shaded, and the finger residue Phe 102 from ABH2 is labelled in green.

## METHODS SUMMARY

The disulphide cross-linked protein–DNA complexes were purified using Mono-Q anion exchange chromatography, and crystals were grown by hanging-drop vapour diffusion. Diffraction data were collected from cryo-preserved crystals at beamlines 19BM and 14BM at the Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois. The structures of AlkB–dsDNA complexes were solved by molecular replacement, and the ABH2–dsDNA structure was solved by MAD phasing with a bromine-labelled DNA. Data collection and refinement parameters for all structures are given in Supplementary Tables 1 and 2. Detailed procedures are presented in Supplementary Information.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 17 December 2007; accepted 5 March 2008.**

1. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* Ch. 1 (ASM Press, Washington DC, 1995).
2. Wood, R. D., Mitchell, M., Sgouros, J. & Lindahl, T. Human DNA repair genes. *Science* **291**, 1284–1289 (2001).
3. Sedgwick, B. Repairing DNA-methylation damage. *Nature Rev. Mol. Cell Biol.* **5**, 148–157 (2004).
4. Mishina, Y. & He, C. Oxidative dealkylation DNA repair mediated by the monomuclear non-heme iron AlkB proteins. *J. Inorg. Biochem.* **100**, 670–678 (2006).
5. Lindahl, T., Sedgewick, B., Sekiguchi, M. & Nakabeppu, Y. Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* **57**, 133–157 (1988).
6. Falnes, P. O., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli. Nature* **419**, 178–182 (2002).
7. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174–178 (2002).
8. Aravind, L. & Koonin, E. V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**, research0007 (2001).
9. Koivisto, P., Robins, P., Lindahl, T. & Sedgwick, B. Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J. Biol. Chem.* **279**, 40470–40474 (2004).
10. Falnes, P. O. Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res.* **32**, 6260–6267 (2004).
11. Delaney, J. C. & Essigmann, J. M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in *alkB Escherichia coli. Proc. Natl Acad. Sci. USA* **101**, 14051–14056 (2004).
12. Mishina, Y., Yang, C.-G. & He, C. Direct repair of the exocyclic DNA adduct 1,$N^6$-ethenoadenine by the DNA repair AlkB proteins. *J. Am. Chem. Soc.* **127**, 14594–14595 (2005).
13. Delaney, J. C. *et al.* AlkB reverses etheno DNA lesions caused by lipid oxidation *in vitro* and *in vivo*. *Nature Struct. Mol. Biol.* **12**, 855–860 (2005).
14. Duncan, T. *et al.* Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl Acad. Sci. USA* **99**, 16660–16665 (2002).
15. Aas, P. A. *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**, 859–863 (2003).
16. Lee, D.-H. *et al.* Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *J. Biol. Chem.* **280**, 39448–39459 (2005).
17. Gerken, T. *et al.* The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469–1472 (2007).
18. Ringvoll, J. *et al.* Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *EMBO J.* **25**, 2189–2198 (2006).
19. Yu, B. *et al.* Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **439**, 879–884 (2006).
20. Sundheim, O. *et al.* Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO J.* **25**, 3389–3397 (2006).

21. Mishina, Y., Chen, L. X. & He, C. Preparation and characterization of the native iron(II)-containing DNA repair AlkB protein directly from *Escherichia coli. J. Am. Chem. Soc.* **126,** 16930–16936 (2004).
22. Dinglay, S., Trewick, S. C., Lindahl, T. & Sedgwick, B. Defective processing of methylated single-stranded DNA by *E. coli* AlkB mutants. *Genes Dev.* **14,** 2097–2105 (2000).
23. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase at 2.7 Å resolution: implications of comformational changes for polymerization and inhibition mechanism. *Science* **282,** 1669–1675 (1998).
24. Verdine, G. L. & Norman, D. P. G. Covalent trapping of protein–DNA complexes. *Annu. Rev. Biochem.* **72,** 337–366 (2003).
25. Mishina, Y. & He, C. Probing the structure and function of the *Escherichia coli* DNA alkylation repair AlkB protein through chemical cross-linking. *J. Am. Chem. Soc.* **125,** 8730–8731 (2003).
26. Duguid, E. M., Mishina, Y. & He, C. How do DNA repair proteins locate potential base lesions? A chemical crosslinking method to investigate the damage-searching mechanism of $O^6$-alkylguanine-DNA alkyltransferases. *Chem. Biol.* **10,** 827–835 (2003).
27. Mishina, Y., Lee, C. H. & He, C. Interaction of human and bacterial AlkB proteins with DNA as probed through chemical cross-linking studies. *Nucleic Acids Res.* **32,** 1548–1554 (2004).
28. Schofield, C. J. & Zhang, Z. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9,** 722–731 (1999).
29. Lange, S. J. & Que, L. Jr. Oxygen activating nonheme iron enzymes. *Curr. Opin. Chem. Biol.* **2,** 159–172 (1998).
30. Elkins, J. M. *et al.* X-ray crystal structure of *Escherichia coli* taurine/α-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* **41,** 5185–5192 (2002).
31. Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanismf for nucleotide flipping and base excision. *Cell* **95,** 249–258 (1998).
32. Fromme, J. C., Banerjee, A., Huang, S. J. & Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427,** 652–656 (2004).
33. Banerjee, A., Santos, W. L. & Verdine, G. L. Structure of a DNA glycocylase searching for lesions. *Science* **311,** 1153–1157 (2006).
34. Slupphaug, G. *et al.* A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384,** 87–92 (1996).
35. Bruner, S. D., Norman, D. P. G. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403,** 859–866 (2000).
36. Parker, J. B. *et al.* Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **449,** 433–437 (2007).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank: R. Zhang and other beamline staff for assistance with data collection; Y. Luo and the Proteomics and Informatics Services Facility at Research Resources Center (University of Illinois at Chicago) for liquid chromatography–mass spectrometry analysis; and X. Yang, H. Chen and P. R. Chen for discussions. We also thank T. Lindahl and B. Sedgwick for the gift of the *abh2* gene. This work was supported by National Institutes of Health (GM071440 to C.H. and a PCBio fellowship for C.T.S.), the W. M. Keck Foundation Distinguished Young Scholar in Medical Research Program (C.H.), and the Arnold and Mabel Beckman Foundation Young Investigator Program (C.H.). Data collection was performed at beamlines 19BM (Structure Biology Center) and 14BM (BioCARS) at the Advanced Photon Source at Argonne National Laboratory; financial support for these beamlines comes from the National Institutes of Health and the United States Department of Energy.

**Author Contributions** C.-G.Y. and C.Y. solved all AlkB–dsDNA and ABH2–dsDNA structures with help from E.M.D. (crystallography), C.T.S. (initial construct of ABH2 and crystallography) and X.J. (biochemistry). P.A.R. contributed to protein crystallography. C.H. designed the overall project and wrote the manuscript with C.-G.Y. and C.Y. All authors discussed results and commented on the manuscript.

**Author Information** Atomic coordinates are deposited in Protein Data Bank under accession numbers 3BKZ ((Mn/2KG) AlkB–DNA1), 3BI3 ((Mn/2KG) AlkB–DNA5), 3BIE ((Mn/2KG)AlkB–DNA4), 3BTX (ABH2–DNA2), 3BTZ (ABH2–DNA3), 3BU0 ((Mn/2KG)ABH2–DNA2), 3BTY (ABH2–DNA6) and 3BUC ((Mn/2KG)ABH2–DNA6). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.H. (chuanhe@uchicago.edu).

## METHODS

**Oligonucleotide synthesis.** Oligonucleotides containing disulphide-tethered cytosine were prepared by incorporating the O4-triazolyl-dU-CE phosphoramidite (Glen Research, Inc.) at the desired position during solid-phase synthesis[25]. All synthetic oligonucleotides were purified with denaturing polyacrylamide gel electrophoresis or reverse-phase high-performance liquid chromatography.

**Cross-linking, purification and crystallization of the AlkB–dsDNA complexes.** A truncated AlkB with deletion of the amino (N)-terminal 11 amino acids[19] was cloned into a pET30a vector (Novagen) and overexpressed in *E. coli* BL21(DE3). The protein was purified following a previously described procedure[21]. The Asp 135 to Cys mutation was introduced using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Synthetic oligonucleotides (DNA1: 5′-TAGGTAAC*ATCGT-3′ and its complement 5′-AACGATATTACCT-3′, where C* is a disulphide-tethered cytosine) were prepared, purified and cross-linked to the D135C mutant AlkB as described[25]. To cross-link away from the active site, a similar procedure was applied to the S129C mutant of AlkB and DNA5 (5′-TAGGTAA(1-meA)AC*CGT and its complement 5′-AACGGTATTACCT-3′ (the 1-meA phosphoramidite from ChemGenes, Inc.)) and DNA4 (5′-TAGGTAA(1-meA)AC*CGT paired with 5′-AACGGTT⌴TACCT-3′, where '⌴' stands for an abasic site). The covalently linked AlkB–dsDNA complexes were purified using Mono-Q anion exchange chromatography (GE Healthcare), after which the buffer was exchanged to 100 mM NaCl, 10 mM Tris-HCl (pH 7.4), and the complexes were concentrated to 12–15 mg μl$^{-1}$. The complexes were crystallized by the hanging drop method at 4 °C. $MnCl_2$ (1.0 mM) and 2KG (2.0 mM) were added to the complexes, and they were mixed in a 1:1 ratio with well solution containing 100 mM NaCl, 25 mM $MgCl_2$, 100 mM cacodylate (pH 6.5) and 20% w/v PEG 8K. Crystals appeared after 2–4 days and were allowed to grow for several weeks. They were transferred into a cryo-protectant solution containing the reservoir solution plus 20% glycerol, and frozen in liquid nitrogen for X-ray data collection.

**Cloning, expression and purification of ABH2-ΔN55.** Limited protease digestion with trypsin and sequence alignment of ABH2 with *E. coli* AlkB and ABH3 led us to design a new construct of ABH2 lacking the first 55 N-terminal amino acids. The abh2-ΔN55 gene was cloned between the *Nde*I and *Hind*III sites of a pET28a+ vector (Novagen). This plasmid was transformed into BL21 (DE3) E. coli cells (Stratagene) for protein overexpression. The protein was purified using Ni-NTA chromatography (GE Healthcare) with an elution buffer of 50 mM sodium phosphate (pH 8.0), 300 mM NaCl and 400 mM imidazole. The N-terminal His-tag was removed by an overnight thrombin digestion at 4 °C and the tag-free protein was purified the next day with another round of Ni-NTA chromatography. About 4–6 mg of protein could be obtained from 1 l of bacterial cells. All other mutants of ABH2-ΔN55 used in this study were generated by using QuikChange Site-Directed Mutagenesis Kit and expressed and purified by the same method.

**Crystallization of the ABH2–dsDNA complexes.** Three additional mutations, C67S, C165S and C192S, were introduced into ABH2-ΔN55 E175C to avoid non-specific disulphide cross-linking. The mutant protein was used for cross-linking with synthetic oligonucleotides DNA2 (5′-CTGTATC*ATTGCG-3′ paired with 5′-TCGCAATAATACA-3′) for the complex that gave the high-resolution structure, and DNA3 (5′-AGGTGAC*AATGCG-3′ paired with 5′-TCGCATTATCACC-3′) for the complex that afforded the low-resolution structure. An ABH2-ΔN55 G169C mutant was cross-linked to DNA6 (5′-CTGTAT(1-meA)AC*TGCG-3′ paired with 5′-TCGCAGTTATACA-3′) to yield the complex that afforded the structure of a disulphide cross-link away from the active site. The covalently linked ABH2–dsDNA complexes were purified using Mono-Q anion exchange chromatography, after which the buffer was exchanged to 100 mM NaCl, 10 mM Tris-HCl (pH 8.0), and the complexes concentrated to 5 mg ml$^{-1}$. Hexagonal rod-shaped crystals grew in 2–4 weeks at 4 °C, in hanging drops containing 1 μl of complex solution and 1 μl of reservoir solution of 100 mM NaCl, 50 mM $MgCl_2$ and 100 mM cacodylate (pH 6.5) and 8% 8K PEG. Crystals were transferred to a cryoprotectant solution of reservoir solution plus 24% glycerol, and were frozen in liquid nitrogen for data collection. Cofactor-containing ABH2–DNA2 crystals grew in 2–4 days at room temperature, in hanging drops containing 1 μl of complex solution (5.0 mg ml$^{-1}$, 100 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1.0 mM $MnCl_2$, 2.0 mM 2KG) and 1 μl of reservoir solution of 100 mM NaCl, 50 mM $MgCl_2$ and 100 mM cacodylate (pH 6.1) and 16% PEG 4K. $MnCl_2$ (2.5 mM) and 2KG were soaked into the ABH2–DNA6 crystals for 4.5 h to give (Mn/2KG)ABH2–DNA6 crystals.

**Structure determination and refinement.** The AlkB–DNA complex structure was phased by molecular replacement (using EPMR[37] and Phaser[38]), using the previously published AlkB structure as a search model. The model was built by using COOT[39] and refined with the program REFMAC5 from the CCP4 suite[40]. To obtain phase information for the ABH2–DNA2 complex structure, 5-BrU-containing oligonucleotide (DNA2: 5′-CTGXATC*AXXGCG-3′ annealed to 5′-TCGCAAXAAXACA-3′, where X stands for 5-Br dU) was used to grow crystals with ABH2. A two-wavelength anomalous diffraction data set was collected at wavelengths of 0.9192 Å (peak) and 0.9199 Å (inflection) at beamline 19BM at the Structure Biology Center at the Applied Photon Source. All X-ray data were integrated and scaled using HKL2000 program suite[41] and converted to structure factors with the CCP4 program. The positions of five bromine atoms from the MAD data were located and the initial electron density map was calculated using the program SOLVE[42]. Refmac5 was used to refine the model against the 2-Å resolution native data. An initial model was built into the electron-density map manually in COOT. The structure was completed using iterative rounds of manual model building in COOT and computational refinement in either CNS[43] or CCP4. Phases for all other ABH2 data sets were determined by molecular replacement using Phaser[38]. Data collection and refinement parameters for all structures are given in Supplementary Tables 1 and 2. Molecular graphics figures were prepared with PyMOL[44].


37. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **55**, 484–491 (1999).
38. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. Erratum. *Acta Crystallogr. D* **59**, 404 (2003).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
40. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D.* **50**, 760–763 (1994).
41. Otwinowski, Z. W. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
42. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
43. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
44. DeLano, W. L. The PyMOL molecular graphics system (DeLano Scientific, Palo Alto, California, 2002).

Vol 452 | 24 April 2008 | doi:10.1038/nature06895

LETTERS

# The inner jet of an active galactic nucleus as revealed by a radio-to-γ-ray outburst

Alan P. Marscher[1], Svetlana G. Jorstad[1,2], Francesca D. D'Arcangelo[1], Paul S. Smith[3], G. Grant Williams[4], Valeri M. Larionov[2], Haruki Oh[1]†, Alice R. Olmstead[1], Margo F. Aller[5], Hugh D. Aller[5], Ian M. McHardy[6], Anne Lähteenmäki[7], Merja Tornikoski[7], Esko Valtaoja[8,9], Vladimir A. Hagen-Thorn[2], Eugenia N. Kopatskaya[2], Walter K. Gear[10], Gino Tosti[11], Omar Kurtanidze[12], Maria Nikolashvili[12], Lorand Sigua[12], H. Richard Miller[13] & Wesley T. Ryle[13]

**Blazars are the most extreme active galactic nuclei. They possess oppositely directed plasma jets emanating at near light speeds from accreting supermassive black holes. According to theoretical models, such jets are propelled by magnetic fields twisted by differential rotation of the black hole's accretion disk or inertial-frame-dragging ergosphere[1–3]. The flow velocity increases outward along the jet in an acceleration and collimation zone containing a coiled magnetic field[4,5]. Detailed observations of outbursts of electromagnetic radiation, for which blazars are famous, can potentially probe the zone. It has hitherto not been possible to either specify the location of the outbursts or verify the general picture of jet formation. Here we report sequences of high-resolution radio images and optical polarization measurements of the blazar BL Lacertae. The data reveal a bright feature in the jet that causes a double flare of radiation from optical frequencies to TeV γ-ray energies, as well as a delayed outburst at radio wavelengths. We conclude that the event starts in a region with a helical magnetic field that we identify with the acceleration and collimation zone predicted by the theories. The feature brightens again when it crosses a standing shock wave corresponding to the bright 'core' seen on the images.**

The jet of BL Lac (Fig. 1) approaches us within 6–10° of our line of sight at a flow speed of 0.981–0.994$c$ (where $c$ is the speed of light), corresponding to a Lorentz factor of $7.0 \pm 1.8$ (ref. 6). Relativistic aberration and the Doppler effect strongly beam the radiation, resulting in the apparent luminosity being hundreds of times higher than it would be if the emitting plasma were at rest. An essentially identical counterjet is presumably present, but too faint to detect because of

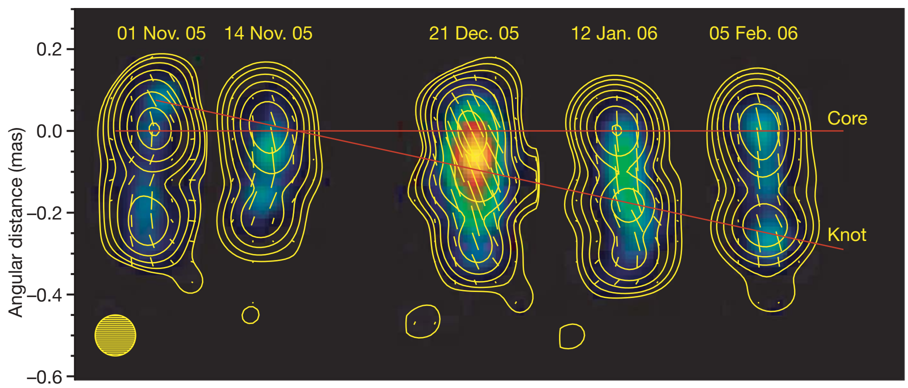



**Figure 1 | Sequence of Very Long Baseline Array images of BL Lac at a wavelength of 7 mm (and a frequency of 43 GHz).** The images are convolved with a circular gaussian function (shown in the bottom-left corner) that has a full width at half maximum of 0.1 mas, the resolution of the longest baselines of the array, which corresponds to 0.12 pc at the distance of 291 Mpc derived from the redshift, $z = 0.069$, and the Hubble Law with an assumed slope of $H_0 = 72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (ref. 22). The vertical axis shows angular distance relative to the core. The contours indicate total flux intensity, starting at 0.023 Janskys (Jy) per beam ($1\,\mathrm{Jy} = 10^{-23}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}}$) and increasing in factors of two. The colour scale represents polarized intensity, and peaks (yellow) at 0.215 Jy per beam. The yellow line segments indicate the electric vector position angle (EVPA) of the polarization, which has an uncertainty of $\pm 7°$ in each image. The EVPA has been rotated by 16° in the clockwise direction to compensate for Faraday rotation; see the Supplementary Information. The approximate path of the centroid of the knot, the definition of which includes a point from an image obtained on 12 March 2006 in addition to those shown, is given by the slanted red line. The knot's observed proper motion of 1.2 mas yr$^{-1}$ is equivalent to an apparent speed of 5.0$c$ after correction for time dilation due to cosmological redshift.

[1]Institute for Astrophysical Research, Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA. [2]Astronomical Institute of St Petersburg State University, Universitetskij Prospect 28, Petrodvorets, 198504 St Petersburg, Russia. [3]Steward Observatory, [4]MMT Observatory, University of Arizona, Tucson, Arizona 85721-0065, USA. [5]Astronomy Department, University of Michigan, 830 Dennison Building, Ann Arbor, Michigan 48109-1090, USA. [6]Department of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK. [7]Metsähovi Radio Observatory, Helsinki University of Technology TKK, Metsähovintie 114, 02540 Kylmälä, Finland. [8]Tuorla Observatory Väisäläntie 20, 21500 Piikkiö, Finland. [9]Department of Physics, University of Turku, 20014 Turku, Finland. [10]School of Physics and Astronomy, Cardiff University, Cardiff CF24 3YB, UK. [11]Department of Physics, University of Perugia, Via A. Pascoli, 06123 Perugia, Italy. [12]Abastumani Astrophysical Observatory, Mt Kanobili, Abastumani, Georgia. [13]Department of Physics and Astronomy, Georgia State University, Atlanta, Georgia 30303, USA. †Present address: Department of Physics, University of California, Berkeley, California 94720-7300, USA.

beaming in the opposite direction. The stationary core lies at the northern end of the visible jet. Bright 'knots' emerge from the core at a rate of 1–2 per year and move south at apparent superluminal speeds, an illusion caused by their relativistic motion[6].

The radio, optical, and X-ray light curves in Fig. 2 indicate a double flare in late 2005. The highly significant detection[7] of $>0.2$ TeV $\gamma$-rays from 2005.819 to 2005.831 during the first X-ray flare implies that acceleration of electrons with sub-TeV energies was particularly efficient at this time. These electrons can both produce X-rays from synchrotron radiation and scatter the X-ray photons to GeV $\gamma$-ray energies that are boosted to the TeV range by relativistic motion of the jet plasma. The location of such flares has been controversial: some observations[8,9] indicate that they occur downstream of the core, whereas most theoretical models require that they take place well upstream of this region, where the plasma is more compact. As we explain below, our data indicate that the first flare in late 2005 corresponds to a disturbance passing through the zone upstream of the core, where the jet flow is still accelerating, and that the second occurs as the disturbance crosses a standing shock system in the core.

The identification of the location of the initial flare within the acceleration and collimation zone is significant, since previous observations of jet collimation are quite limited. For example, an image[10] at 7-mm wavelength of the radio galaxy M87 appears to reveal an initially broad outflow that narrows into a nearly cylindrical jet. This is consistent with gradual collimation by either a toroidal magnetic field[4] or external confining gas pressure that decreases with distance from the black hole[11]. The flow seen in M87 could include a 'sheath' that moves more slowly and is less focused than the 'spine'[12]. In the case of BL Lac, the high apparent superluminal motions of bright knots in the jet and the pronounced variability at all wavelengths imply that the observed radiation arises exclusively from the spine, where special relativistic effects dominate.

The primary observational indicator of magnetic collimation requiring a coiled magnetic field in the spine of the jet is the evolution of the polarization. When observed at an angle to its axis, synchrotron radiation from a circularly symmetric jet with a helical field displays a net polarization oriented either parallel or perpendicular to the projected jet axis[13]. Such parallel and perpendicular polarizations can be confused with shock waves and velocity shear, respectively, which can produce the same polarization patterns. However, in a model where magnetic forces gradually accelerate and focus the jet, the flow velocity is directed along streamlines that follow a helical trajectory with a different, wider, pitch angle than that of the magnetic field[5]. The rotation of the flow can be traced back to the base of the jet in the orbiting accretion disk or differentially rotating ergosphere, where the spin of the black hole drags the inertial frames. A shock wave or other compressive feature propagating down the jet traces a spiral path that follows a streamline and cycles through the orientations of the helical field (see Fig. 3 and ref. 5). This should manifest itself as a rotation of the position angle of linear polarization as the feature moves outward. The degree of polarization should drop to a minimum in the middle of the rotation, when the mean magnetic field in the flaring region is transverse to that of the previously existing emission[14]. As Fig. 2g, h demonstrates, we see both effects.

The optical EVPA shown in Fig. 2g rotates steadily by about 240° over a five-day interval before settling at a value of ~195°. The

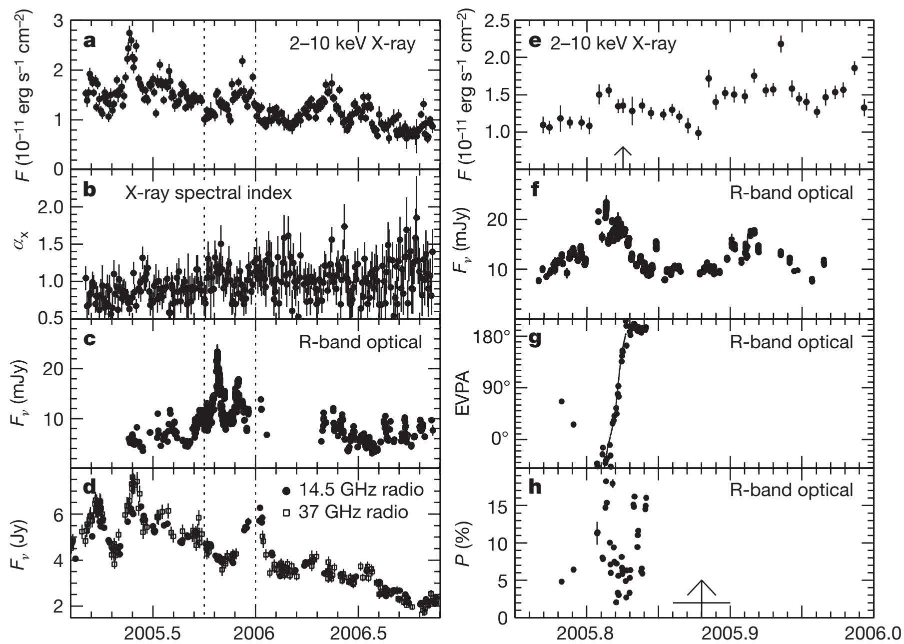



**Figure 2 | Flux density at various wavebands and optical polarization of BL Lac, as functions of time. a–d**, Dependence on time of the flux of radiation from BL Lac over a two-year interval at the indicated wavebands. The X-ray measurements in **a** are of photon energy flux $F$ integrated over photon energies of 2–10 keV. Error bars represent $\pm 1$ s.d. uncertainties in the values plotted. The exponent of the power-law dependence of X-ray flux density on frequency is denoted by $-\alpha_x$. **e–h**, Enlargements of the 0.25-yr time interval marked by vertical dotted lines in panels **a–d**, but with optical R-band EVPA (**g**) and degree of polarization $P$ (**h**) respectively replacing X-ray spectral index (**b**) and radio flux density (**d**) (whereas **e** and **f** respectively show the magnified intervals in **a** and **c**). Error bars represent $\pm 1$ s.d. The interval of highly significant detections[7] at photon energies $>0.2$ TeV is indicated by the width of the head of the arrow in **e**. The rotation in optical R-band EVPA near the time of the peak of the first optical and X-ray flare is apparent. Because there is an ambiguity of $\pm 180°$ in the value of the EVPA, we have selected the quadrant of each value that provides a consistent overall trend of rotation between 2005.81 and 2005.83. The solid curve in **g** corresponds to the pattern predicted by the model shown in Fig. 3 when relativistic aberration is included. The vertical arrow (with error bar) in **h** indicates the time at which the superluminal knot is coincident with the stationary core seen in the images displayed in Fig. 1. Optical polarimetric data were obtained from Steward Observatory and the Crimean Astrophysical Observatory. Optical flux density points were obtained from photometry at these two sites plus Lowell Observatory, Perugia University Astronomical Observatory and the Abastumani Astrophysical Observatory. All of the optical telescopes are equipped with charge-coupled-device cameras. Measurements of X-ray flux and the continuum spectrum were obtained from a monitoring program with the NASA Rossi X-ray Timing Explorer. Measurements of radio flux density were obtained from the University of Michigan Radio Astronomy Observatory and the Metsähovi Radio Observatory. Descriptions of telescopes and data analysis are available in the Supplementary Information.

sequence of images in Fig. 1 shows a bright, superluminal knot that first appears upstream of the core. It subsequently moves past the core and proceeds down the jet at a position angle of $\sim 190^\circ$ and with an EVPA that is parallel to the jet to within the observational uncertainty. The close correspondence between the optical and 7-mm EVPAs after 29 October implies that the knot is the emitter of the polarized optical emission during the flare.

Previous authors[14–16] have suggested that rotations of the polarization vector occur in BL Lac and the similar active galactic nucleus OJ 287. Their observations, which were more poorly sampled than ours, allowed multiple interpretations owing to the $\pm 180^\circ$ ambiguity of the EVPA. Despite this, the model that we advocate is quite similar to one of those proposed previously[14,15], with the location of the emission region and connection with high-energy flares now specified by our sequences of Very Long Baseline Array images and multiwaveband light curves.

We interpret the event in the following manner (see Fig. 3). Explosive activity at the inlet of the jet near the black hole injects a surge of energy into the jet across part of its cross-sectional area. This disturbance appears as a knot of emission as it propagates along a subset of streamlines through the acceleration and collimation zone. Doppler beaming of radiation emitted by high-energy electrons in the disturbance increases as the knot accelerates along its spiral path, which stretches out with distance down the jet. These effects cause the flux of synchrotron radiation from the knot to rise until it dominates the optical, X-ray and (through inverse Compton scattering) $\gamma$-ray emission from BL Lac as the disturbance exits the zone of helical magnetic field. Maximum beaming—and therefore the peak in the light curve of the first flare—occurs during the last spiral, when the Lorentz factor of the jet is near its asymptotic value and the velocity vector of the knot points most closely towards our line of sight. The peak can be quite sharp[5], as observed. At the point when the flare dominates the optical flux, we see the optical polarization vector rotate before the shock exits the acceleration and collimation zone. This zone is opaque at radio wavelengths, owing to synchrotron self-absorption; hence, the first flare is absent in the radio light curves.

Beyond the acceleration and collimation zone, the disturbance forms a moving shock wave that encounters a region of turbulence, which is possibly driven by velocity shear across the jet[6] downstream of the point at which the magnetic and particle energy densities reach rough equipartition[4]. The ambient magnetic field in the jet has a chaotic structure in this region. Because the shock front amplifies only the component of the field that is parallel to the front, the EVPA becomes transverse to this direction and therefore essentially

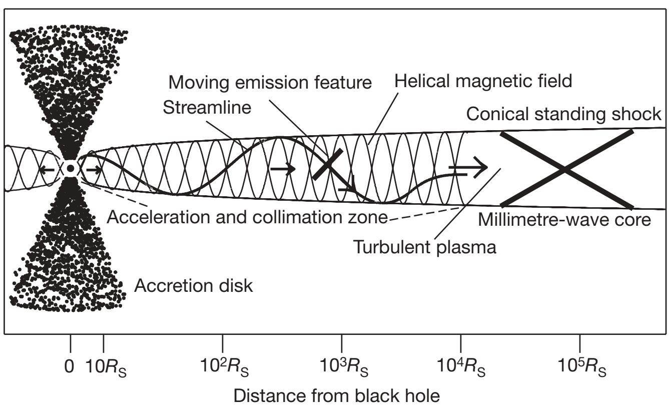



**Figure 3 | Proposed model for the inner jet of BL Lac.** A shock propagates down the jet along a spiral streamline. The first flare occurs during the last $240^\circ$ twist of the streamline before the flow straightens and becomes turbulent. The passage of the feature through the millimetre-wave core stimulates the second flare. A logarithmic scale of distance from the black hole, shown in terms of the Schwarzschild radius ($R_S$), is used to illustrate phenomena on various scales.

parallel to the velocity vector of the knot at a position angle of $\sim 190^\circ$. During this phase, the flux lessens as the knot continues down the broadening jet, where the magnetic field strength and electron density decrease.

According to the model we propose here, the variation of EVPA with time should deviate from a strict linear dependence owing to projection effects, because the circular cross-section has an elliptical shape from our viewpoint. We have calculated this effect, including relativistic aberration, and show in Fig. 2g that the optical EVPA data do follow the predicted curve. The small number of brief excursions of the EVPA from the curve, the deviations from the mean EVPA before and after the rotation, and irregularities in the light curves can all be explained by local flare-ups of emission that briefly amplify both the polarization along a particular direction and the flux at various wavebands.

The smoothness of the plot of EVPA versus time eliminates the possibility[16,17] that the rotation is actually a random walk of the polarization vector due to a chaotic magnetic field. If this were the case, our numerical simulations (see ref. 17) indicate that the curve would be much more jagged than is observed when the degree of polarization is $\sim 5\%$. In the simulations, this level of polarization corresponds to synchrotron emission from $\sim 200$ independent cells, each with a randomly oriented magnetic field. Apparent rotations by $\sim 240^\circ$ are very rare in such simulations, whereas they are common during flares of BL Lac and similar objects[14].

Both synchrotron radiation and inverse Compton scattering contribute to the X-ray emission from BL Lac, with synchrotron radiation dominating when electrons are accelerated to energies in the TeV range[18,19]. This generally causes the plot of flux density ($F_\nu$) versus frequency ($\nu$) to steepen in the X-ray range in such a way that the spectral index $\alpha$ is greater than one, where $F_\nu \propto \nu^{-\alpha}$. Such X-ray spectral steepening occurs during the first flare. In contrast, the X-ray spectrum becomes harder ($\alpha < 1$) during the second flare, as is expected if the X-rays are mostly generated by inverse Compton scattering of optical and infrared photons.

The second flare, which started at 2005.89, is simultaneous with the passage of the knot through the core seen on the Very Long Baseline Array images. If the core were a standing conical shock, as has been determined from simultaneous radio and optical polarization variability in the case of the quasar PKS 0420−014 (ref. 18), the emission would increase as the knot undergoes compression by the shock front. The flare dies down at optical and X-ray frequencies as the knot propagates away from the core down the expanding jet. However, it lasts much longer at 43 GHz, at which frequency the synchrotron radiation requires lower-energy electrons that have longer energetic lifetimes than those emitting at higher frequencies.

In the Supplementary Information we relate the angular velocity of the feature, inferred from the rotation of the optical polarization vector, to the rotational velocity of the base of the magnetic field near the black hole. We find that the rotational velocity thus obtained is consistent with the predictions of models in which the jet is driven by twisting magnetic fields from either the accretion disk[1,3,4,5] or the ergosphere of the black hole[2,3,20,21].

The combination of densely sampled sub-milliarcsecond imaging using the Very Long Baseline Array, polarimetry, and multiwaveband flux measurements has allowed us to explore the inner jet of BL Lac. Future data from more sensitive $\gamma$-ray Cherenkov detectors and the NASA Gamma-ray Large Area Space Telescope will soon allow more refined analyses. The ability to detect emission upstream of the core at 7-mm wavelength indicates that increasing the resolution of very-long-baseline interferometry by adding antennas in high Earth orbits will provide more detailed direct imaging of the inner jets of active galactic nuclei.

**Received 17 January; accepted 6 March 2008.**

1. Blandford, R. D. & Payne, D. G. Hydromagnetic flows from accretion discs and the production of radio jets. *Mon. Not. R. Astron. Soc.* **199,** 883–903 (1982).

2. Blandford, R. D. & Znajek, R. Electromagnetic extraction of energy from Kerr black holes. *Mon. Not. R. Astron. Soc.* **179,** 433–456 (1977).
3. Meier, D. L., Koide, S. & Uchida, Y. Magnetohydrodynamic production of relativistic jets. *Science* **291,** 84–92 (2001).
4. Vlahakis, N. & Königl, A. Magnetic driving of relativistic outflows in active galactic nuclei. I. Interpretation of parsec-scale accelerations. *Astrophys. J.* **605,** 656–661 (2004).
5. Vlahakis, N. in *Blazar Variability Workshop II: Entering the GLAST Era (ASP Conf. Ser. 350)* (eds Miller, H. R., Marshall, K., Webb, J. R. & Aller, M. F.) 169–177 (Astronomical Society of the Pacific, San Francisco, 2006).
6. Jorstad, S. G. *et al.* Polarimetric observations of 15 active galactic nuclei at high frequencies: jet kinematics from bimonthly monitoring with the Very Long Baseline Array. *Astron. J.* **130,** 1418–1466 (2005).
7. Albert, J. *et al.* Discovery of very high energy $\gamma$-ray emission from the low-frequency peaked BL Lac object BL Lacertae. *Astrophys. J.* **666,** L17–L20 (2007).
8. Jorstad, S. G. *et al.* Multi-epoch VLBA observations of EGRET-detected quasars and BL Lac objects: connection between superluminal ejections and gamma-ray flares in Blazars. *Astrophys. J.* **556,** 738–748 (2001).
9. Lähteenmäki, A. & Valtaoja, E. Testing of inverse Compton models for active galactic nuclei with gamma-ray and radio observations. *Astrophys. J.* **590,** 95–108 (2003).
10. Junor, W., Biretta, J. A. & Livio, M. Formation of the radio jet in M87 at 100 Schwarzschild radii from the central black hole. *Nature* **401,** 891–892 (1999).
11. Melia, F., Liu, S. & Fatuzzo, F. Is thermal expansion driving the initial gas ejection in NGC 6251? *Astrophys. J.* **567,** 811–816 (2002).
12. Laing, R. A., Parma, P., de Ruiter, H. R. & Fanti, R. Asymmetries in the jets of weak radio galaxies. *Mon. Not. R. Astron. Soc.* **306,** 513–530 (1999).
13. Lyutikov, M., Pariev, V. I. & Gabuzda, D. C. Polarisation and structure of relativistic parsec-scale AGN jets. *Mon. Not. R. Astron. Soc.* **360,** 869–891 (2005).
14. Sillanpää, A., Takalo, L. O., Nilsson, K. & Kikuchi, S. Photopolarimetry of BL Lac. *Astrophys. Space Sci.* **206,** 55–70 (1993).
15. Moore, R. L. *et al.* The noise of BL Lacertae. *Astrophys. J.* **260,** 415–436 (1982).
16. Kikuchi, S., Inoue, M., Mikami, Y., Tabara, H. & Kato, T. A synchronous variation of polarization angle in OJ 287 in the optical and radio regions. *Astron. Astrophys.* **190,** L8–L10 (1988).
17. D'Arcangelo, F. D. *et al.* Rapid multiwaveband polarization variability in the quasar PKS 0420–014: optical emission from the compact radio jet. *Astrophys. J.* **659,** L107–L110 (2007).
18. Böttcher, M. *et al.* Coordinated multiwavelength observations of BL Lacertae in 2000. *Astrophys. J.* **596,** 847–859 (2003).
19. Tanihata, C. *et al.* Rapid synchrotron flares from BL Lacertae detected by ASCA and RXTE. *Astrophys. J.* **543,** 124–130 (2000).
20. McKinney, J. C. & Narayan, R. Disk-jet coupling in black hole accretion systems II: force-free electrodynamical models. *Mon. Not. R. Astron. Soc.* **375,** 531–547 (2007).
21. Hawley, J. & Krolik, J. H. Magnetically driven jets in the Kerr metric. *Astrophys. J.* **641,** 103–116 (2006).
22. Freedman, W. L. *et al.* Final results from the Hubble Space Telescope key project to measure the Hubble constant. *Astrophys. J.* **553,** 47–72 (2001).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** The authors thank N. Vlahakis for discussions. The research reported here is based on work supported in part by the US National Science Foundation, NASA, the Russian Fund for Basic Research, the Academy of Finland, and the Italian Space Agency (ASI). The Very Long Baseline Array is an instrument of the National Radio Astronomy Observatory, a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.P.M. (marscher@bu.edu).

Vol 452|24 April 2008|doi:10.1038/nature06843

# LETTERS

# A topological Dirac insulator in a quantum spin Hall phase

D. Hsieh[1], D. Qian[1], L. Wray[1], Y. Xia[1], Y. S. Hor[2], R. J. Cava[2] & M. Z. Hasan[1,3]

**When electrons are subject to a large external magnetic field, the conventional charge quantum Hall effect[1,2] dictates that an electronic excitation gap is generated in the sample bulk, but metallic conduction is permitted at the boundary. Recent theoretical models suggest that certain bulk insulators with large spin–orbit interactions may also naturally support conducting topological boundary states in the quantum limit[3–5], which opens up the possibility for studying unusual quantum Hall-like phenomena in zero external magnetic fields[6]. Bulk $Bi_{1-x}Sb_x$ single crystals are predicted to be prime candidates[7,8] for one such unusual Hall phase of matter known as the topological insulator[9–11]. The hallmark of a topological insulator is the existence of metallic surface states that are higher-dimensional analogues of the edge states that characterize a quantum spin Hall insulator[3–13]. In addition to its interesting boundary states, the bulk of $Bi_{1-x}Sb_x$ is predicted to exhibit three-dimensional Dirac particles[14–17], another topic of heightened current interest following the new findings in two-dimensional graphene[18–20] and charge quantum Hall fractionalization observed in pure bismuth[21]. However, despite numerous transport and magnetic measurements on the $Bi_{1-x}Sb_x$ family since the 1960s[17], no direct evidence of either topological Hall states or bulk Dirac particles has been found. Here, using incident-photon-energy-modulated angle-resolved photoemission spectroscopy (IPEM-ARPES), we report the direct observation of massive Dirac particles in the bulk of $Bi_{0.9}Sb_{0.1}$, locate the Kramers points at the sample's boundary and provide a comprehensive mapping of the Dirac insulator's gapless surface electron bands. These findings taken together suggest that the observed surface state on the boundary of the bulk insulator is a realization of the 'topological metal'[9–11]. They also suggest that this material has potential application in developing next-generation quantum computing devices that may incorporate 'light-like' bulk carriers and spin-textured surface currents.**

Bismuth is a semimetal with strong spin–orbit interactions. Its band structure is believed to feature an indirect negative gap between the valence band maximum at the T-point of the bulk Brillouin zone and the conduction band minima at three equivalent L-points[17,22] (here we generally refer to these as a single point, L). The valence and conduction bands at L are derived from antisymmetric ($L_a$) and symmetric ($L_s$) *p*-type orbitals, respectively, and the effective low-energy hamiltonian at this point is described by the (3+1)-dimensional relativistic Dirac equation[14–16]. The resulting dispersion relation, $E(\mathbf{k}) = \pm\sqrt{(\mathbf{v}\cdot\mathbf{k})^2 + \Delta^2} \approx \mathbf{v}\cdot\mathbf{k}$, is highly linear owing to the combination of an unusually large band velocity $\mathbf{v}$ and a small gap $\Delta$ (such that $|\Delta/|\mathbf{v}|| \approx 5 \times 10^{-3}$ Å$^{-1}$) and has been used to explain various peculiar properties of bismuth[14–16]. Substituting bismuth with antimony is believed to change the critical energies of the band structure as follows (see Fig. 1). At an Sb concentration of $x \approx 4\%$, the gap $\Delta$ between $L_a$ and $L_s$ closes and a massless, three-dimensional (3D) Dirac point is realized. As $x$ is further increased this gap re-opens with inverted symmetry ordering, which leads to a change in sign of $\Delta$ at each of the three equivalent L-points in the Brillouin zone. For concentrations greater than $x \approx 7\%$ there is no overlap between the valence band at T and the conduction band at L, and the material becomes an inverted-band insulator. Once the band at T drops below the valence band at L, at $x \approx 8\%$, the system evolves into a direct-gap insulator whose low-energy physics is dominated by the spin–orbit-coupled Dirac particles at L[7,17].

Recently, semiconductors with inverted bandgaps have been proposed to manifest the two-dimensional (2D) quantum spin Hall phase, which is predicted to be characterized by the presence of metallic one-dimensional edge states[3–5,12]. Although a band-inversion mechanism and edge states have been invoked to interpret the transport results in 2D mercury telluride semiconductor quantum wells[13], no one-dimensional edge states are directly imaged, so their topological character is yet to be uniquely determined. Recent theoretical treatments[7,8] have focused on the strongly spin–orbit-coupled, band-inverted $Bi_{1-x}Sb_x$ series as a possible 3D bulk realization of the quantum spin Hall phase in which the one-dimensional edge states are expected to take the form of 2D surface states[7–9] that may be directly imaged and spectroscopically studied, making it feasible to identify their topological order parameter character.

High-momentum-resolution angle-resolved photoemission spectroscopy performed with varying incident photon energy (IPEM-ARPES) allows for the measurement of electronic band dispersion along various momentum space (**k**-space) trajectories in the 3D bulk Brillouin zone. ARPES spectra taken along two orthogonal cuts through the L-point of the bulk Brillouin zone of $Bi_{0.9}Sb_{0.1}$ are shown in Fig. 1a, c. A Λ-shaped dispersion whose tip lies less than 50 meV below the Fermi energy $E_F$ can be seen along both directions. Additional features originating from surface states that do not disperse with incident photon energy are also seen. Owing to the finite intensity between the bulk and surface states, the exact binding energy $E_B$ where the tip of the Λ-shaped dispersion lies is unresolved. The linearity of the bulk Λ-shaped bands is observed by locating the peak positions at higher $E_B$ in the momentum distribution curves and the energy at which these peaks merge is obtained by extrapolating linear fits to the momentum distribution curves. Therefore, 50 meV represents a lower bound on the energy gap $\Delta$ between $L_a$ and $L_s$. The magnitudes of the extracted band velocities along the $k_x$ and $k_y$ directions are $7.9 \pm 0.5 \times 10^4$ m s$^{-1}$ and $10.0 \pm 0.5 \times 10^5$ m s$^{-1}$, respectively, which are similar to the tight binding values of $7.6 \times 10^4$ m s$^{-1}$ and $9.1 \times 10^5$ m s$^{-1}$ calculated for the $L_a$ band of bismuth[22]. Our data are consistent with the extremely small effective mass of $0.002m_e$ (where $m_e$ is the electron mass) observed in magneto-reflection measurements on samples with $x = 11\%$ (ref. 23). The Dirac point in

[1]Joseph Henry Laboratories of Physics, Department of Physics, [2]Department of Chemistry, [3]Princeton Center for Complex Materials, Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, New Jersey 08544, USA.

graphene, coincidentally, has a band velocity ($|\mathbf{v}_F| \approx 10^6\,\mathrm{m\,s^{-1}}$)[18] comparable to what we observe for $Bi_{0.9}Sb_{0.1}$, but its spin–orbit coupling is several orders of magnitude weaker, and the only known method of inducing a gap in the Dirac spectrum of graphene is by coupling to an external chemical substrate[20]. The $Bi_{1-x}Sb_x$ series thus provides a rare opportunity to study relativistic Dirac hamiltonian physics in a 3D condensed matter system where the intrinsic (rest) mass gap can easily be tuned.

Studying the band dispersion perpendicular to the sample surface provides a way to differentiate bulk states from surface states in a 3D material. To visualize the near-$E_F$ dispersion along the 3D L–X cut (X is a point that is displaced from L by a $k_z$-distance of $3\pi/c$, where $c$ is the lattice constant), in Fig. 2a we plot energy distribution curves (EDCs) taken such that electrons at $E_F$ have fixed in-plane momentum ($k_x$, $k_y$) = ($L_x$, $L_y$) = (0.8 Å$^{-1}$, 0.0 Å$^{-1}$), as a function of photon energy $h\nu$. There are three prominent features in the EDCs: a non-dispersing, $k_z$-independent peak centred just below $E_F$, at about −0.02 eV; a broad, non-dispersing hump centred near −0.3 eV; and a strongly dispersing hump that coincides with the latter near $h\nu$ = 29 eV. To understand which bands these features originate from, we show ARPES intensity maps along an in-plane cut $\bar{K}\bar{M}\bar{K}$ (parallel to the $k_y$-direction) taken using $h\nu$ values of 22 eV, 29 eV and 35 eV, which correspond to approximate $k_z$ values of $L_z$ − 0.3 Å$^{-1}$, $L_z$, and $L_z$ + 0.3 Å$^{-1}$, respectively (Fig. 2b). At $h\nu$ = 29 eV, the low-energy ARPES spectral weight reveals a clear Λ-shaped band close to $E_F$. As the photon energy is either increased or decreased from 29 eV, this intensity shifts to higher binding energies as the spectral weight evolves from the Λ-shaped band into a 'U'-shaped band. Therefore, the dispersive peak in Fig. 2a comes from the bulk valence band, and for $h\nu$ = 29 eV the high-symmetry point L = (0.8, 0, 2.9) appears in the third bulk Brillouin zone. In the maps of Fig. 2b with respective $h\nu$ values of 22 eV and 35 eV, overall weak features near $E_F$ that vary in intensity remain even as the bulk valence band moves far below $E_F$. The survival of these weak features over a large photon energy range (17–55 eV) supports their surface origin. The non-dispersing feature centred near −0.3 eV in Fig. 2a comes from the

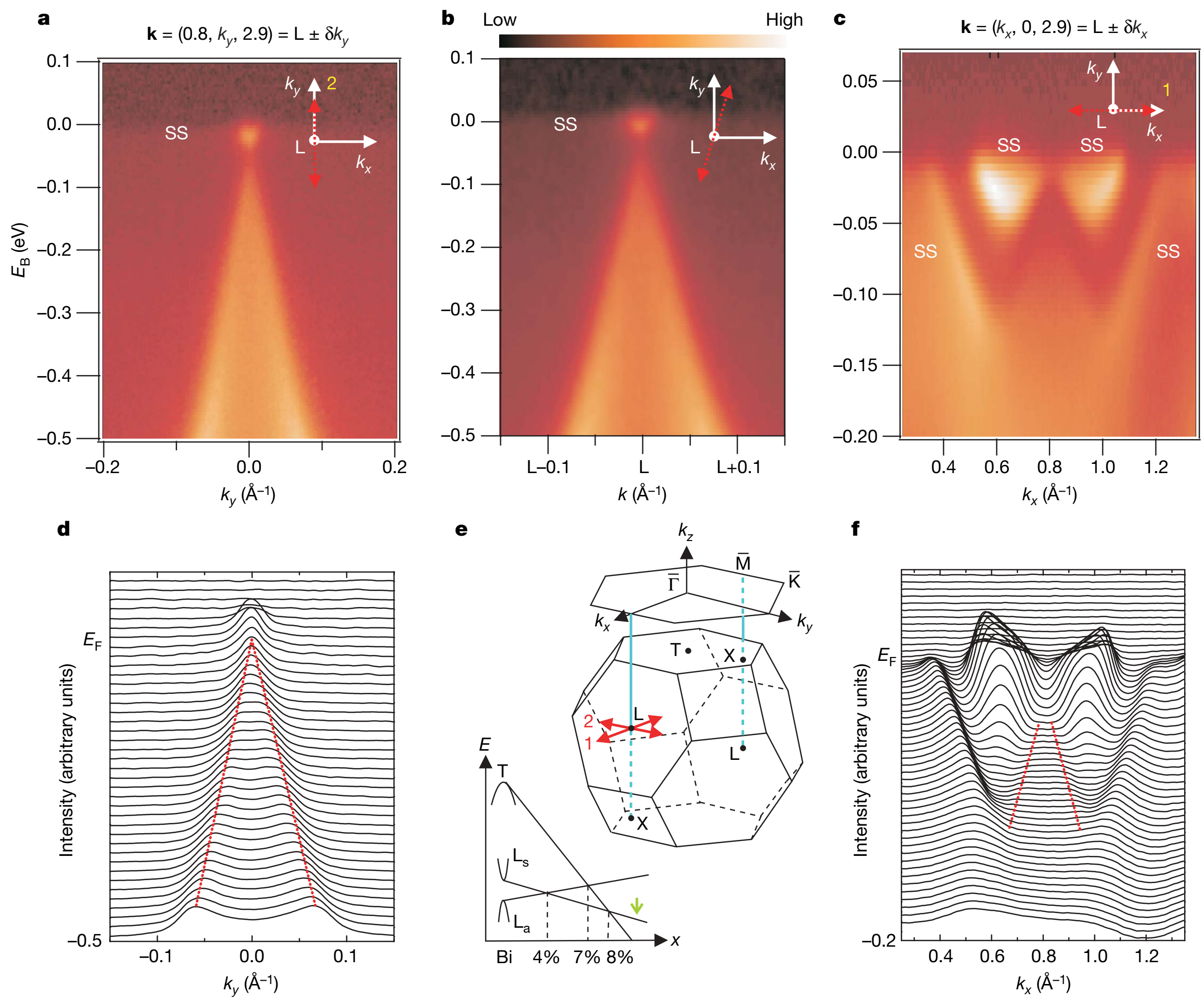



**Figure 1 | Dirac-like dispersion near the L-point in the bulk Brillouin zone.** Selected ARPES intensity maps of $Bi_{0.9}Sb_{0.1}$ are shown along three **k**-space cuts through the L-point of the bulk 3D Brillouin zone. The presented data are taken in the third Brillouin zone with $L_z$ = 2.9 Å$^{-1}$ with a photon energy of 29 eV. The cuts are along the $k_y$-direction (**a**); a direction rotated by approximately 10° from the $k_y$-direction (**b**); and the $k_x$-direction (**c**). Here, δ symbolizes a change along a particular **k**-direction. Each cut shows a Λ-shaped bulk band whose tip lies below the Fermi level, signalling a bulk gap. The surface states are denoted SS and are all identified in Fig. 2 (for further identification via theoretical calculations, see Supplementary Information). **d**, Momentum distribution curves corresponding to the intensity map in **a**. **f**, A log-scale plot of the momentum distribution curves corresponding to the intensity map in **c**. The red lines are guides to the eye for the bulk features in the momentum distribution curves. **e**, Schematic of the bulk 3D Brillouin zone of $Bi_{1-x}Sb_x$ and the 2D Brillouin zone of the projected (111) surface. The high-symmetry points $\bar{\Gamma}$, $\bar{M}$ and $\bar{K}$ of the surface Brillouin zone are labelled. The schematic evolution of bulk band energies as a function of $x$ is shown. The L-band inversion transition occurs at $x \approx 0.04$, where a 3D gapless Dirac point is realized, and the composition we study here (for which $x$ = 0.1) is indicated by the green arrow. A more detailed phase diagram based on our experiments is shown in Fig. 3c.

higher-binding-energy (valence band) part of the full spectrum of surface states, and the weak non-dispersing peak at −0.02 eV reflects the low-energy part of the surface states that cross $E_F$ away from the point $\bar{M}$ and forms the surface Fermi surface (Fig. 2c).

Having established the existence of an energy gap in the bulk state of $Bi_{0.9}Sb_{0.1}$ (Figs 1 and 2) and observed linearly dispersive bulk bands uniquely consistent with strong spin–orbit coupling model calculations[14–16,22] (see Supplementary Information for full comparison with the theoretical calculation), we now discuss the topological character of its surface states, which we found to be gapless (Fig. 2c). In general, the states at the surface of spin–orbit-coupled compounds are allowed to be spin-split owing to the loss of space inversion symmetry ($E(k,\uparrow) = E(-k,\uparrow)$). However, as required by Kramers' theorem, this splitting must go to zero at the four time-reversal-invariant momenta in the 2D surface Brillouin zone. As discussed in refs 7 and 9, along a path connecting two time-reversal-invariant momenta in the same Brillouin zone, the Fermi energy inside the bulk gap will intersect these singly degenerate surface states either an even or an odd number of times. When there are an even number of surface state crossings, the surface states are topologically trivial because weak disorder (as may arise through alloying) or correlations can remove pairs of such crossings by pushing the surface bands entirely above or below $E_F$. When there are an odd number of crossings, however, at least one surface state must remain gapless, which makes it non-trivial[7–9]. The existence of such topologically non-trivial surface states can be theoretically predicted on the basis of the bulk band structure only, using the $Z_2$-invariant that is related to the quantum Hall Chern number[24]. Materials with band structures with $Z_2 = +1$ are ordinary Bloch band insulators that are topologically equivalent to the filled-shell atomic insulator, and are predicted to exhibit an even number (including zero) of surface state crossings. Materials with bulk band structures with $Z_2 = -1$, on the other hand, which are expected to exist in rare systems with strong spin–orbit coupling acting as an internal quantizing magnetic field

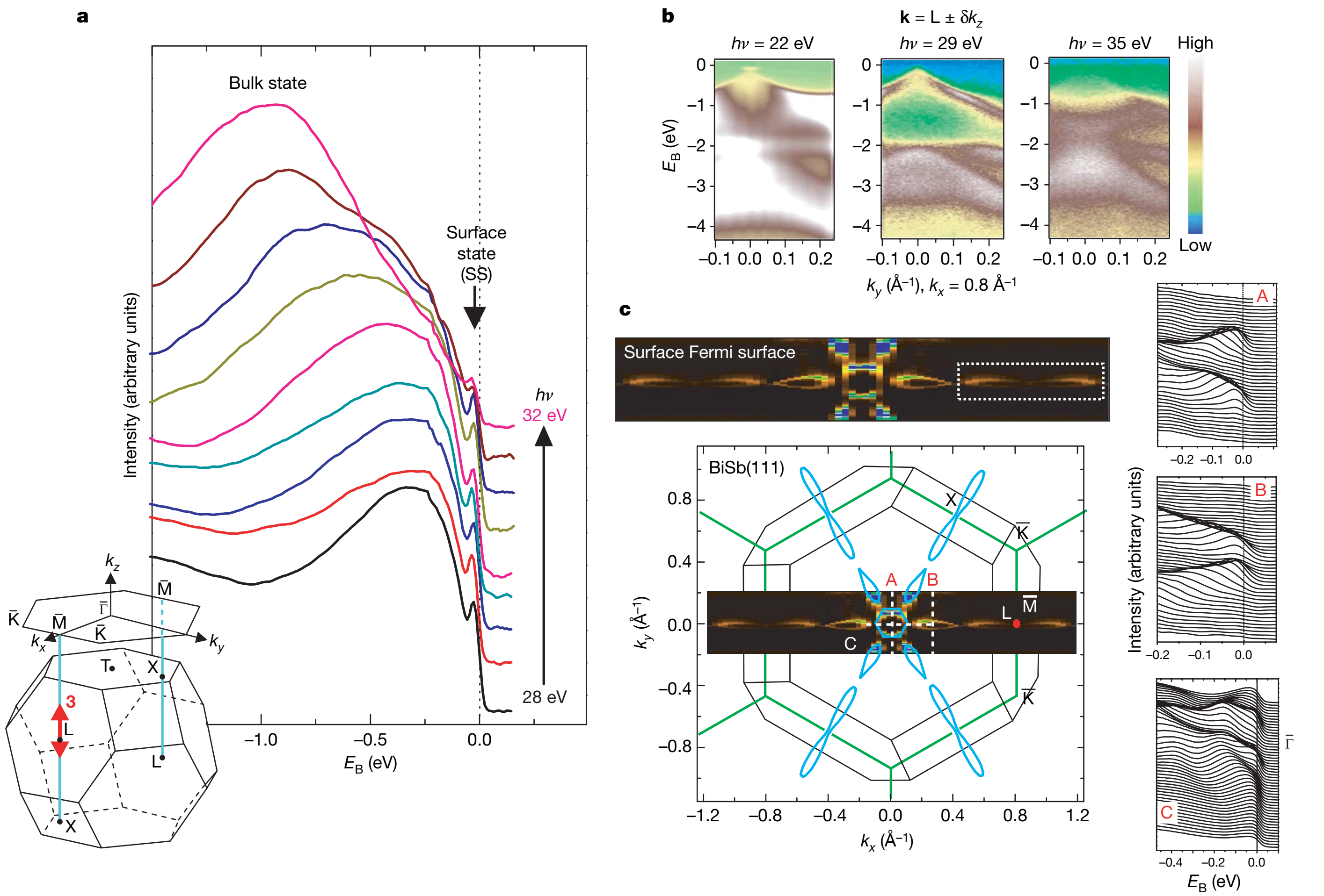



**Figure 2 | Dispersion along the cut in the $k_z$-direction.** Surface states are experimentally identified by studying their out-of-plane momentum dispersion through the systematic variation of incident photon energy. **a**, EDCs of $Bi_{0.9}Sb_{0.1}$ with electrons at the Fermi level ($E_F$) maintained at a fixed in-plane momentum of ($k_x = 0.8$ Å$^{-1}$, $k_y = 0.0$ Å$^{-1}$) are obtained as a function of incident photon energy to identify states that exhibit no dispersion perpendicular to the (111) plane along the direction shown by the double-headed arrow labelled '3' in the inset (see Methods for the detailed procedure). Selected EDC data sets with photon energies ranging from 28 eV to 32 eV in steps of 0.5 eV are shown for clarity. The non-energy-dispersive ($k_z$-independent) peaks near $E_F$ are the surface states. **b**, ARPES intensity maps along cuts parallel to the $k_y$-direction taken with electrons at $E_F$ fixed at $k_x = 0.8$ Å$^{-1}$ with respective photon energies of $h\nu = 22$ eV, $h\nu = 29$ eV and $h\nu = 35$ eV (for a conversion map from photon energy to $k_z$, see Supplementary Information). The faint Λ-shaped band at $h\nu = 22$ eV and $h\nu = 35$ eV show some overlap with the bulk valence band at L ($h\nu = 29$ eV), suggesting that it is a resonant surface state degenerate with the bulk state in some limited range of momenta near $E_F$. The flat band of intensity centred about −2 eV in the $h\nu = 22$ eV scan originates from Bi 5$d$ core-level emission from second-order light. **c**, Projection of the bulk Brillouin zone (black lines) onto the (111) surface Brillouin zone (green lines). The overlay (enlarged in inset) shows the high-resolution Fermi surface of the metallic surface state mode, which was obtained by integrating the ARPES intensity (taken with $h\nu = 20$ eV) from −15 meV to 10 meV with respect to $E_F$. The six teardrop-shaped lobes of the Fermi surface close to $\bar{\Gamma}$ (the centre of the Brillouin zone) show some intensity variation between them that is due to the relative orientations between the axes of the lobes and the axis of the detector slit. The six-fold symmetry was however confirmed by rotating the sample in the $k_x$–$k_y$ plane. EDCs corresponding to the cuts A, B and C are also shown; these confirm the gapless character of the surface states in bulk insulating $Bi_{0.9}Sb_{0.1}$.

on the electron system[6] and inverted bands at an odd number of high-symmetry points in their bulk 3D Brillouin zones, are predicted to exhibit an odd number of surface state crossings, precluding their adiabatic continuation to the atomic insulator[3,7–13]. Such 'topological metals'[9–11] cannot be realized in a purely 2D electron gas system.

In our experimental case, namely the (111) surface of $Bi_{0.9}Sb_{0.1}$, the four time-reversal-invariant momenta are located at $\bar{\Gamma}$ and three $\bar{M}$-points that are rotated by 60° relative to one another. Owing to the three-fold crystal symmetry (A7 bulk structure) and the observed mirror symmetry of the surface Fermi surface across $k_x = 0$ (Fig. 2), these three $\bar{M}$-points are equivalent (and we henceforth refer to them as a single point, $\bar{M}$). The mirror symmetry ($E(k_y) = E(-k_y)$) is also expected, from time-reversal invariance exhibited by the system. The complete details of the surface state dispersion observed in our experiments along a path connecting $\bar{\Gamma}$ and $\bar{M}$ are shown in Fig. 3a; finding this information is made possible by our experimental separation of surface states from bulk states. As for bismuth, two surface bands emerge from the bulk band continuum near $\bar{\Gamma}$ to form a central electron pocket and an adjacent hole lobe[25–27]. It has been established that these two bands result from the spin-splitting of a surface state and are thus singly degenerate[27,28].

On the other hand, the surface band that crosses $E_F$ at $-k_x \approx 0.5$ Å$^{-1}$, and forms the narrow electron pocket around $\bar{M}$, is clearly doubly degenerate, as far as we can determine within our experimental resolution. This is indicated by its splitting below $E_F$ between $-k_x \approx 0.55$ Å$^{-1}$ and $\bar{M}$, as well as the fact that this splitting goes to zero at $\bar{M}$ in accordance with Kramers' theorem. In semimetallic single-crystal bismuth, only a single surface band is observed to form the electron pocket around $\bar{M}$ (refs 29 and 30). Moreover, this surface state overlaps, and hence becomes degenerate with, the bulk conduction band at L (L projects to the surface point $\bar{M}$) owing to the semimetallic character of bismuth (Fig. 3b). In $Bi_{0.9}Sb_{0.1}$, on the other hand, the states near $\bar{M}$ fall completely inside the bulk energy gap, preserving their purely surface character at $\bar{M}$ (Fig. 3a). The surface

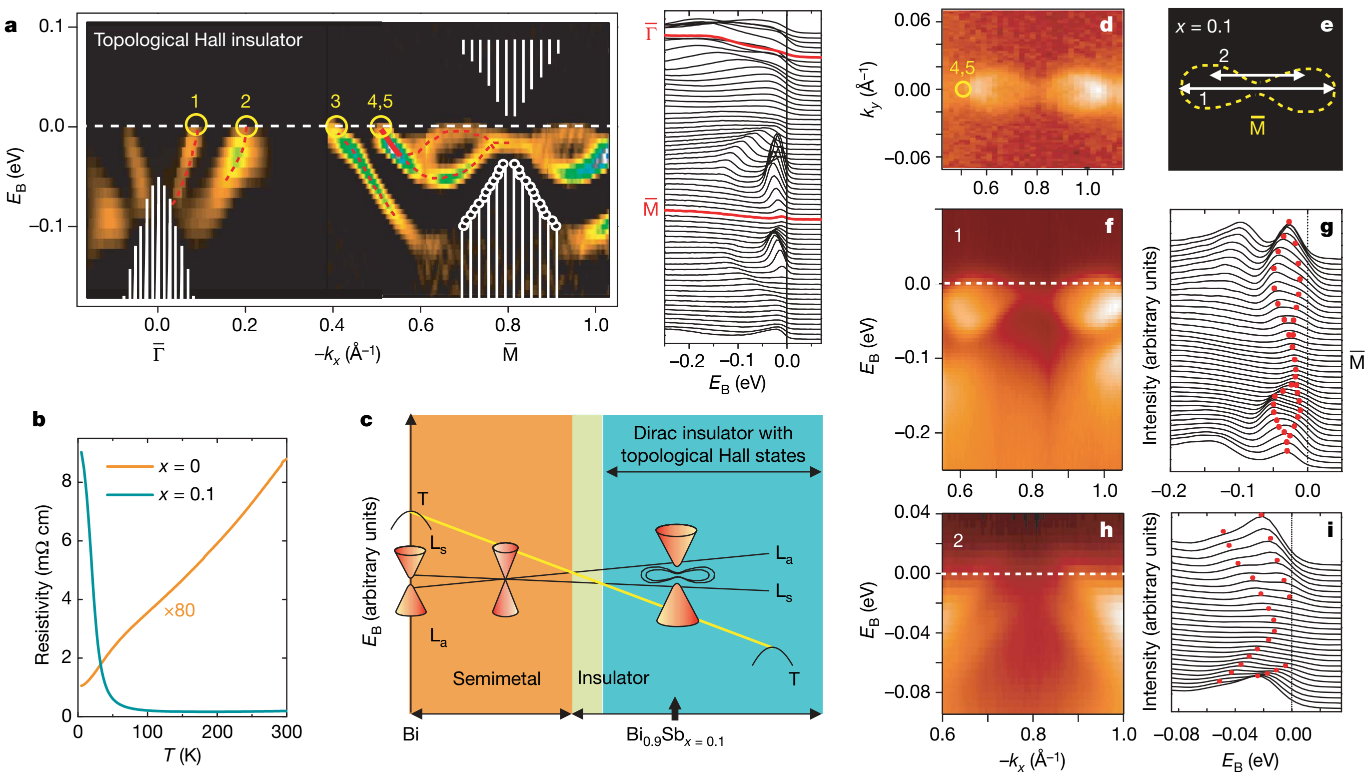



**Figure 3 | The topological gapless surface states in bulk insulating $Bi_{0.9}Sb_{0.1}$.** **a**, The surface-band-dispersion second-derivative image of $Bi_{0.9}Sb_{0.1}$ along $\bar{\Gamma}-\bar{M}$. The shaded white area shows the projection of the bulk bands based on ARPES data, as well as a rigid shift of the tight binding bands to sketch the unoccupied bands above the Fermi level. A non-intrinsic flat band of intensity near $E_F$ generated by analysis of the second-derivative image was rejected to isolate the intrinsic dispersion. The Fermi crossings of the surface state are denoted by yellow circles, with the band near $-k_x \approx 0.5$ Å$^{-1}$ counted twice owing to double degeneracy. The red lines are guides to the eye. An in-plane rotation of the sample by 60° produced the same surface state dispersion. The EDCs along $\bar{\Gamma}-\bar{M}$ are shown in the right-hand diagram. There are a total of five crossings between $\bar{\Gamma}$ and $\bar{M}$, which indicates that these surface states are topologically non-trivial. The number of surface state crossings in a material (with an odd number of Dirac points) is related to the topological $Z_2$ invariant (see text). **b**, The resistivity curves of Bi and $Bi_{0.9}Sb_{0.1}$ reflect the contrasting transport behaviours. The presented resistivity curve for pure bismuth has been multiplied by a factor of 80 for clarity. **c**, Schematic variation of bulk band energies of $Bi_{1-x}Sb_x$ as a function of $x$ (based on band calculations and on refs 7 and 17). $Bi_{0.9}Sb_{0.1}$ is a direct-gap bulk Dirac point insulator well inside the inverted-band regime, and its surface forms a 'topological metal'—the 2D analogue of the one-dimensional edge states in quantum spin Hall systems. **d**, ARPES intensity integrated within ±10 meV of $E_F$ originating solely from the surface state crossings. The image was plotted by stacking along the negative $k_x$-direction a series of scans taken parallel to the $k_y$-direction. **e**, Outline of the $Bi_{0.9}Sb_{0.1}$ surface state ARPES intensity near $E_F$ measured in **d**. White lines show the scan directions '1' and '2'. **f**, Surface band dispersion along direction '1' taken with $h\nu = 28$ eV, and the corresponding EDCs (**g**). The surface Kramers degenerate point, critical in determining the topological $Z_2$ class of a band insulator, is clearly seen at $\bar{M}$, approximately 15 ± 5 meV below $E_F$. (We note that the scans are taken along the negative $k_x$-direction, away from the bulk L-point.) **h**, Surface band dispersion along direction '2' taken with $h\nu = 28$ eV, and the corresponding EDCs (**i**). This scan no longer passes through the $\bar{M}$-point, and the observation of two well-separated bands indicates the absence of Kramers degeneracy as expected, which corroborates the result in **a**.

Kramers doublet point can thus be defined in the bulk insulator (unlike in bismuth[25–30]) and is experimentally located in $Bi_{0.9}Sb_{0.1}$ samples to lie approximately 15 ± 5 meV below $E_F$ at $\mathbf{k} = \bar{M}$ (Fig. 3a). For the precise location of this Kramers point, it is important to demonstrate that our alignment is strictly along the $\bar{\Gamma}$–$\bar{M}$ line. To do so, we contrast high-resolution ARPES measurements taken along the $\bar{\Gamma}$–$\bar{M}$ line with those that are slightly offset from it (Fig. 3e). Figures 3f–i show that with $k_y$ offset from the Kramers point at $\bar{M}$ by less than 0.02 $\text{Å}^{-1}$, the degeneracy is lifted and only one band crosses $E_F$ to form part of the bow-shaped electron distribution (Fig. 3d). Our finding of five surface state crossings (an odd rather than an even number) between $\bar{\Gamma}$ and $\bar{M}$ (Fig. 3a), confirmed by our observation of the Kramers degenerate point at the time-reversal-invariant momentum, indicates that these gapless surface states are topologically non-trivial. This corroborates our bulk electronic structure result that $Bi_{0.9}Sb_{0.1}$ is in the insulating band-inverted ($Z_2 = -1$) regime (Fig. 3c), which contains an odd number of bulk (gapped) Dirac points, and is topologically analogous to an integer-quantum-spin Hall insulator.

Our experimental results taken collectively strongly suggest that $Bi_{0.9}Sb_{0.1}$ is quite distinct from graphene[18,19] and represents a novel state of quantum matter: a strongly spin–orbit-coupled insulator with an odd number of Dirac points and a negative $Z_2$ topological Hall phase. Our work points to future possibilities for further spectroscopic investigations of topological orders in quantum matter.

## METHODS SUMMARY

High-resolution IPEM-ARPES data have been taken at beamlines 12.0.1 and 10.0.1 of the Advanced Light Source at the Lawrence Berkeley National Laboratory, as well as at the PGM beamline of the Synchrotron Radiation Center in Wisconsin, with photon energies ranging from 17 eV to 55 eV and energy resolutions ranging from 9 meV to 40 meV, and momentum resolution better than 1.5% of the surface Brillouin zone. Data were taken on high-quality bulk single-crystal $Bi_{1-x}Sb_x$ at a temperature of 15 K and chamber pressures less than $8 \times 10^{-11}$ torr. Throughout this paper, the bulk bands presented are from those measured in the third bulk Brillouin zone, to ensure a high degree of signal-to-noise contrast, and the $k_z$ values are estimated using the standard free-electron final-state approximation.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 25 November 2007; accepted 14 February 2008.**

1. von Klitzing, K., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45,** 494–497 (1980).
2. Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48,** 1559–1562 (1982).
3. Kane, C. L. & Mele, E. J. Quantum spin Hall effect in graphene. *Phys. Rev. Lett.* **95,** 226801 (2005).
4. Bernevig, B. A. & Zhang, S.-C. Quantum spin Hall effect. *Phys. Rev. Lett.* **96,** 106802 (2006).
5. Sheng, D. N., Weng, Z. Y., Sheng, L. & Haldane, F. D. M. Quantum spin-Hall effect and topological Chern numbers. *Phys. Rev. Lett.* **97,** 036808 (2006).
6. Haldane, F. D. M. Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the 'parity anomaly'. *Phys. Rev. Lett.* **61,** 2015–2018 (1988).
7. Fu, L. & Kane, C. L. Topological insulators with inversion symmetry. *Phys. Rev. B* **76,** 045302 (2007).
8. Murakami, S. Phase transition between the quantum spin Hall and insulator phases in 3D: emergence of a topological gapless phase. *New J. Phys.* **9,** 356 (2007).
9. Fu, L., Kane, C. L. & Mele, E. J. Topological insulators in three dimensions. *Phys. Rev. Lett.* **98,** 106803 (2007).
10. Moore, J. E. & Balents, L. Topological invariants of time-reversal-invariant band structures. *Phys. Rev. B* **75,** 121306(R) (2007).
11. Roy, R. Three dimensional topological invariants for time reversal invariant Hamiltonians and the three dimensional quantum spin Hall effect. ⟨http://arXiv.org/abs/cond-mat/0607531v1⟩ (2006).
12. Bernevig, B. A., Hughes, T. L. & Zhang, S.-C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314,** 1757–1761 (2006).
13. Konig, M. *et al.* Quantum spin Hall insulator state in HgTe quantum wells. *Science* **318,** 766–770 (2007).
14. Wolff, P. A. Matrix elements and selection rules for the two-band model of bismuth. *J. Phys. Chem. Solids* **25,** 1057–1068 (1964).
15. Fukuyama, H. & Kubo, R. Interband effects of magnetic susceptibility. II. Diamagnetism of bismuth. *J. Phys. Soc. Jpn* **28,** 570–581 (1970).
16. Buot, F. A. Weyl transformation and the magnetic susceptibility of a relativistic Dirac electron gas. *Phys. Rev. A* **8,** 1570–1581 (1973).
17. Lenoir, B. *et al.* in *Fifteenth International Conference on Thermoelectrics (Pasadena, California)* 1–13 (IEEE, New York, 1996).
18. Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438,** 201–204 (2005).
19. Novoselov, K. S. *et al.* Room-temperature quantum Hall effect in graphene. *Science* **315,** 1379 (2007).
20. Zhou, S. Y. *et al.* Substrate-induced bandgap opening in epitaxial graphene. *Nature Mater.* **6,** 770–775 (2007).
21. Behnia, K., Balicas, L. & Kopelevich, Y. Signatures of electron fractionalization in ultraquantum bismuth. *Science* **317,** 1729–1731 (2007).
22. Liu, Y. & Allen, E. Electronic structure of semimetals Bi and Sb. *Phys. Rev. B* **52,** 1566–1577 (1995).
23. Hebel, L. C. & Smith, G. E. Interband transitions and band structure of a BiSb alloy. *Phys. Lett.* **10,** 273–275 (1964).
24. Kane, C. L. & Mele, E. J. $Z_2$ topological order and the quantum spin Hall effect. *Phys. Rev. Lett.* **95,** 146802 (2005).
25. Ast, C. R. & Hochst, H. Fermi surface of Bi(111) measured by photoemission spectroscopy. *Phys. Rev. Lett.* **87,** 177602 (2001).
26. Hochst, H. & Gorovikov, S. Lack of electron-phonon coupling along two-dimensional bands in $Bi_{1-x}Sb_x$ single crystal alloys. *J. Elect. Spectrosc. Relat. Phenom.* **144,** 351–355 (2005).
27. Hofmann, P. The surfaces of bismuth: Structural and electronic properties. *Prog. Surf. Sci.* **81,** 191–245 (2006).
28. Hirahara, T. *et al.* Direct observation of spin splitting in bismuth surface states. *Phys. Rev. B* **76,** 153305 (2007).
29. Hengsberger, M. *et al.* Photoemission study of the carrier bands in Bi(111). *Eur. Phys. J. B* **17,** 603–608 (2000).
30. Ast, C. R. & Hochst, H. Electronic structure of a bismuth bilayer. *Phys. Rev. B* **67,** 113102 (2003).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank P. W. Anderson, B. A. Bernevig, L. Balents, E. Demler, A. Fedorov, F. D. M. Haldane, D. A. Huse, C. L. Kane, R. B. Laughlin, J. E. Moore, N. P. Ong, A. N. Pasupathy, D. C. Tsui and S.-C. Zhang for discussions. The synchrotron experiments are supported by the DOE-BES and materials synthesis is supported by the NSF-MRSEC at Princeton Center for Complex Materials.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.Z.H. (mzhasan@princeton.edu).

## METHODS

**Growth method for high-quality single crystals.** The $Bi_{1-x}Sb_x$ single-crystal samples ($0 \leq x \leq 0.17$) used for ARPES experiments were each cleaved from a boule grown from a stoichiometric mixture of high-purity elements. The boule was cooled from 650 °C to 270 °C over a period of five days and was annealed for seven days at 270 °C. The samples naturally cleaved along the (111) plane, which resulted in shiny flat silver surfaces. X-ray diffraction measurements were used to check that the samples were single phase, and confirmed that the $Bi_{0.9}Sb_{0.1}$ single crystals presented in this paper have a rhombohedral A7 crystal structure (point group $R\bar{3}m$), with room-temperature (300 K) lattice parameters $a = 4.51$ Å and $c = 11.78$ Å indexed using a rhombohedral unit cell. The X-ray diffraction patterns of the cleaved crystals exhibit only the (333), (666), and (999) peaks, showing that the cleaved surface is oriented along the trigonal (111) axis. Room-temperature data were recorded on a Bruker D8 diffractometer using Cu K$\alpha$ radiation ($\lambda = 1.54$ Å) and a diffracted-beam monochromator. The in-plane crystal orientation was determined by Laue X-ray diffraction. During the angle-resolved photoemission spectroscopy (ARPES) measurements a fine alignment was achieved by carefully studying the band dispersions and Fermi surface symmetry as an internal check for crystal orientation.

**Transport measurements.** Temperature-dependent resistivity measurements were carried out on single-crystal samples in a Quantum Design PPMS-9 instrument, using a standard four-probe technique on approximately $4 \times 1 \times 1$-mm$^3$, rectangular samples with the current in the basal plane, which was perpendicular to the trigonal axis. The four contacts were made by using room-temperature silver paste. The data for samples with concentrations ranging from $x = 0$ to $x = 0.17$ showed a systematic change from semimetallic to insulating-like behaviour with increasing $x$, in agreement with previously published works[15], which was used as a further check of the antimony concentrations. Conventional magnetic and transport measurements[7,17,31] such as these cannot separately measure the contributions of the surface and bulk states to the total signal. ARPES, on the other hand, is a momentum-selective technique[32], which allows for a separation of 2D (surface) from 3D (bulk) dispersive energy bands. This capability is especially important for $Bi_{1-x}Sb_x$ because the Dirac point lies at a single point in the 3D Brillouin zone, unlike for 2D graphene, where the Dirac points can be studied at any arbitrary perpendicular momentum along a line[33,34].

**Systematic methods for separating bulk from surface electronic states.** ARPES is a photon-in, electron-out technique[32]. Photoelectrons ejected from a sample by a monochromatic beam of radiation are collected by a detector capable of measuring its kinetic energy $E_{kin}$. By varying the detector angles, $\theta$ and $\varphi$, relative to the sample surface normal, the momentum of the photoelectrons, **K**, can also be determined (as illustrated in Supplementary Fig. 1a). By employing the commonly used free-electron final state approximation, we can fully convert from the measured kinetic energy and momentum values of the photoelectron to the binding energy, $E_B$, and Bloch momentum values **k** of its initial state inside the crystal, via

$$|E_B| = h\nu - W - E_{kin}$$

$$k_x = K_x = \frac{1}{\hbar}\sqrt{2m_e E_{kin}}\sin\theta$$

$$k_z = \frac{1}{\hbar}\sqrt{2m_e(E_{kin}\cos^2\theta - V_0)}$$

where we have set $\varphi = 0$, $W$ is the work function, $m_e$ is the electron mass and $V_0$ is an experimentally determined parameter, which is approximately $-10$ eV for bismuth[35,36]. Features in the ARPES spectra originating from bulk initial states (dispersive along the $k_z$-direction) were distinguished from those originating from surface initial states (non-dispersive along the $k_z$-direction) by studying their dependence on incident photon energy, $h\nu$, and converting this to dependence on $k_z$ via the displayed equations. ARPES data were collected at beamlines 12.0.1 and 10.0.1 of the Advanced Light Source at the Lawrence Berkeley National Laboratory, as well as at the PGM beamline of the Synchrotron Radiation Center in Wisconsin, with incident photon energies ranging from 17 eV to 55 eV, energy resolutions ranging from 9 meV to 40 meV and momentum resolution better than 1.5% of the surface Brillouin zone, using Scienta electron analysers. The combination of high spatial resolution and high crystalline quality enabled us to probe only the highly ordered and cleanest regions of our samples. Single-crystal $Bi_{1-x}Sb_x$ samples were cleaved *in situ* at a temperature of 15 K and chamber pressures less than $8 \times 10^{-11}$ torr, and high surface quality was checked throughout the measurement process by monitoring the EDC linewidths of the surface state. To measure the near-$E_F$ dispersion of an electronic band along a direction normal to the sample surface, such as the direction from X ($2\pi/\sqrt{3}a, 0, 8\pi/c$) to L ($2\pi/\sqrt{3}a, 0, 11\pi/c$) shown in Fig. 2a, EDCs were taken at several incident photon energies. The kinetic energy of the photoelectron at $E_F$ is different for each value of $h\nu$, so the angle $\theta$ was first adjusted and then held fixed for each $h\nu$ so as to keep $k_x$ constant at $2\pi/\sqrt{3}a = 0.8$ Å$^{-1}$ for electrons emitted near $E_F$. To ensure that the in-plane momentum remained constant at $\bar{M}$ (the L–X line projects onto $\bar{M}$) for each EDC, a complete near-$E_F$ intensity map was generated for each photon energy to precisely locate the $\bar{M}$-point (see Supplementary Fig. 1d). We note that because the bulk crystal has only three-fold rotational symmetry about the $k_z$-axis, the reciprocal lattice does not have mirror symmetry about the $k_x = 0$ plane. Therefore, scans taken at $+\theta$ and $-\theta$ for the same photon energy probe different points in the bulk 3D Brillouin zone; this is responsible for the absence of the bulk $\Lambda$-shaped band in Fig. 3f.


31. Kopelevich, Y. *et al.* Universal magnetic-field-driven metal-insulator-metal transformations in graphite and bismuth. *Phys. Rev. B* **73,** 165128 (2006).
32. Hufner, S. *Photoelectron Spectroscopy* (Springer, Berlin, 1995).
33. Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438,** 197–200 (2005).
34. Bostwick, A., Ohta, T., Seyller, T., Horn, K. & Rotenberg, E. Quasiparticle dynamics in graphene. *Nature Phys.* **3,** 36–40 (2007).
35. Jezequel, G., Thomas, J. & Pollini, I. Experimental band structure of semimetal bismuth. *Phys. Rev. B* **56,** 6620–6626 (1997).
36. Ast, C. R. & Hochst, H. High-resolution mapping of the three-dimensional band structure of Bi(111). *Phys. Rev. B* **70,** 245122 (2004).

Vol 452 | 24 April 2008 | doi:10.1038/nature06874




# Small phonon contribution to the photoemission kink in the copper oxide superconductors

Feliciano Giustino[1,2], Marvin L. Cohen[1,2] & Steven G. Louie[1,2]

**Despite over two decades of intense research efforts, the origin of high-temperature superconductivity in the copper oxides remains elusive. Angle-resolved photoemission spectroscopy experiments[1,2] have revealed a kink in the dispersion relations (energy versus wavevector) of electronic states in the copper oxides at binding energies of 50−80 meV, raising the hope that this anomaly could be a key to understanding high-temperature superconductivity. The kink is often interpreted in terms of interactions between the electrons and a bosonic field. Although there is no consensus on the nature of the bosons (or even whether a boson model is appropriate), both phonons[1] and spin fluctuations[2] have been proposed as possible candidates. Here we report first-principles calculations of the role of phonons and the electron–phonon interaction in the photoemission spectra of $La_{2-x}Sr_xCuO_4$. Our calculations within the standard formalism demonstrate that the phonon-induced renormalization of the electron energies and the Fermi velocity is almost one order of magnitude smaller than the effect observed in photoemission experiments. Therefore, our result rules out electron–phonon interaction in bulk $La_{2-x}Sr_xCuO_4$ as the exclusive origin of the measured kink. Our conclusions are consistent with those reached independently in a recent study[3] of the related compound $YBa_2Cu_3O_7$.**

Previous first-principles investigations have already suggested a rather weak electron–phonon coupling[4,5] in $Ca_{0.27}Sr_{0.63}CuO_2$ and $YBa_2Cu_3O_7$. However, because these studies did not directly address the signature of the electron–phonon interaction in the photoemission spectra, their implications for the interpretation of the kink are unclear. In the present study we aim to perform a direct and quantitative comparison between the measured and the calculated photoemission kinks in a prototypical copper oxide superconductor. We studied $La_{2-x}Sr_xCuO_4$ (LSCO) because it crystallizes in a simple lattice characterized by a single $CuO_2$ layer, and the hole concentration can be experimentally controlled over the entire phase diagram by varying the strontium content[6]. We considered optimally doped ($x = 0.15$) and heavily overdoped ($x = 0.30$) LSCO, for which a mean-field description of the electronic structure such as that given by density functional theory is appropriate to interpret measured Fermi surfaces[7]. We did not attempt to address the electron–phonon coupling in the underdoped compound because for that case the validity of the Fermi liquid approach has been questioned[8]. In order to calculate the electronic structure of LSCO we employed a generalized gradient approximation to density functional theory[9,10] (see Supplementary Information). The calculated Fermi surfaces of LSCO are generally in good agreement with angle-resolved photoemission spectroscopy experiments[11] (see Supplementary Fig. 1), and in particular they reproduce the evolution of the Fermi surface topology from a (0, 0)-centred electron pocket in overdoped LSCO to a (π, π)-centred hole pocket in the underdoped compound[12] (wavevectors are given in units of $1/a$, $a = 3.77$ Å being the lattice parameter).

We studied the lattice dynamics of LSCO by calculating the phonon eigenmodes and eigenenergies across the entire Brillouin zone using density functional perturbation theory[13]. In Fig. 1 we compare the calculated phonon dispersions along the antinodal direction (0,0)–(π,0) and the total vibrational density of states to inelastic neutron scattering data[14,15]. The dispersion of the in-plane

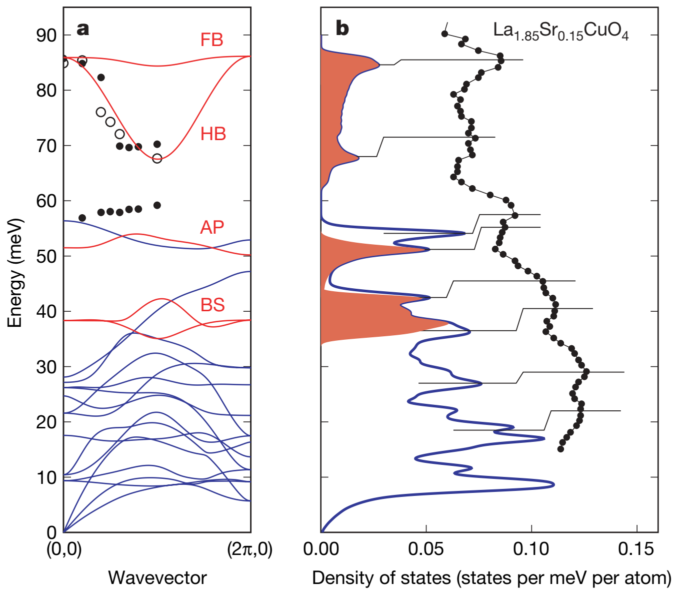



**Figure 1 | Phonons of $La_{2-x}Sr_xCuO_4$ at optimal doping ($x = 0.15$).** **a**, Calculated dispersions along the antinodal direction (0, 0)–(2π, 0) (solid lines), compared with the inelastic neutron scattering data of ref. 14 (circles: 300 K data; disks: 10 K data). The labels indicate the phonon branches associated with the full-breathing (FB) in-plane Cu–O stretching mode at (π, π), the half-breathing (HB) in-plane Cu–O stretching mode at (π, 0), the $c$ axis apical (AP) oxygen stretching mode at (0, 0), and the buckling/stretching (BS) modes at (π, 0) and (π, π). **b**, Calculated phonon density of states (solid line), compared with the inelastic neutron scattering data of ref. 15 (disks). The stepped black lines indicate the correspondence between the measured and calculated peaks. The theoretical phonon energies agree with experiment to within 5 meV. The shaded regions correspond to the phonon branches highlighted in **a**. The calculated half-breathing (π, 0) and apical (0, 0) frequencies respectively correspond to energies of 68 meV and 53 meV, whereas the experimental values are 68–70 meV and 57 meV (ref. 14). To extend the present comparison to the undoped parent oxide, we also performed the calculations for the (π, 0) half-breathing mode in antiferromagnetic $La_2CuO_4$. The theoretical energy of 76 meV agrees with neutron scattering measurements yielding 74 meV (ref. 16), and indicates that our theoretical framework correctly describes the softening of the half-breathing phonon mode through the insulator-to-metal transition[15–17].

[1]Department of Physics, University of California at Berkeley, [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

half-breathing Cu–O stretching mode is found to be in very good agreement with experiment. Inspection of the vibrational density of states shows that a one-to-one correspondence can be established between the calculated and measured peaks, indicating an overall agreement between theory and experiment[14–17].

An electron travelling in a solid distorts the lattice because of the Coulomb interaction with the ions. The lattice distortion in turn has a feedback on the electron dynamics, resulting in an increase of the electron mass and a shortening of the electron lifetime in a particular quasi-particle state. Within quantum field theory, this effect is described in terms of a complex self-energy $\Sigma$ that the electron acquires as a consequence of the electron–phonon interaction. The real part of the self-energy describes the change in the electron energy, and the imaginary part provides information on the electron lifetime $\tau$ through $\tau = \hbar/(2\mathrm{Im}\Sigma)$. The electron self-energy is calculated within the Migdal approximation to Feynman–Dyson perturbation theory[18–20]. Because of the $d$-wave symmetry of the pseudogap[21] and the superconducting gap[22], this is a reasonable approximation for the (0, 0)–($\pi$, $\pi$) nodal direction. By focusing on the nodal direction we also avoid the van Hove singularity at ($\pi$, 0), and the comparison with experiment can be performed with little ambiguity. The diagonal part of the electron self-energy operator $\Sigma(\mathbf{r},\mathbf{r}',\omega)$ with respect to the unperturbed electronic states arising from the electron–phonon interaction is[18–20]

$$\begin{aligned}\Sigma_{n\mathbf{k}}(\omega) &= \langle n\mathbf{k}\,|\,\Sigma(\mathbf{r},\mathbf{r}',\omega)\,|\,n\mathbf{k}\rangle \\ &= \sum_{m\nu}\int \frac{\mathrm{d}\mathbf{k}'}{\Omega_{\mathrm{BZ}}}\,|g_{mn,\nu}(\mathbf{k},\mathbf{k}')|^2 \\ &\quad\times\left[\frac{1-f_{m\mathbf{k}'}+n_{\mathbf{q}\nu}}{\omega-\varepsilon_{m\mathbf{k}'}-\omega_{\mathbf{q}\nu}-i\delta}+\frac{f_{m\mathbf{k}'}+n_{\mathbf{q}\nu}}{\omega-\varepsilon_{m\mathbf{k}'}+\omega_{\mathbf{q}\nu}-i\delta}\right]\end{aligned} \qquad (1)$$

where $\mathbf{r}$ and $\mathbf{r}'$ correspond to position vectors; $\omega$ is the frequency of the quasi-particle excitation; $|n\mathbf{k}\rangle$ denotes a Bloch eigenstate with wavevector $\mathbf{k}$, band index $n$ and energy $\varepsilon_{n\mathbf{k}}$; and $g_{mn,\nu}(\mathbf{k},\mathbf{k}') = \langle m\mathbf{k}'|\,\Delta_{\mathbf{q}\nu}V(\mathbf{r})\,|n\mathbf{k}\rangle$ is the electron–phonon matrix element for the scattering $|n\mathbf{k}\rangle \rightarrow |m\mathbf{k}'\rangle$ through a phonon of wavevector $\mathbf{q} = \mathbf{k}' - \mathbf{k}$, branch index $\nu$ and energy $\omega_{\mathbf{q}\nu}$. The operator $\Delta_{\mathbf{q}\nu}V(\mathbf{r})$ represents the variation of the self-consistent potential with respect to a collective lattice displacement associated with this phonon, and $f_{m\mathbf{k}'}$ and $n_{\mathbf{q}\nu}$, which are respectively the fermion and boson occupation numbers, account for the temperature dependence of the self-energy.

The accurate evaluation of the electron self-energy requires the calculation of the electron–phonon matrix elements $g_{mn,\nu}(\mathbf{k},\mathbf{k}')$ on an extremely fine mesh consisting of more than 100,000 phonon wavevectors. This represents a formidable computational task and was previously beyond the reach of existing approaches. To overcome this difficulty we developed a new technique that is based on the Wannier representation[23] and exploits the short-range nature of the electron–phonon interaction in real space[24,25]. Figure 2 shows the electron self-energy calculated for optimally doped and overdoped LSCO at 20 K. At optimal doping the real part of the self-energy displays a peak at a binding energy of 70 meV, a shoulder at a binding energy of 40 meV and a broad background extending up to much higher binding energies. A mode-resolved analysis shows that the main peak at 70 meV can be assigned to the Cu–O in-plane half-breathing motion around ($\pi$, 0) and full-breathing motion around ($\pi$, $\pi$). The shoulder at 40 meV arises from phonons involving components of both out-of-plane buckling of the planar oxygen atoms and in-plane O–O stretching vibrations. The broad high-energy background can be shown to arise from a density-of-states effect (see Fig. 2 legend). As a consequence of the larger hole concentration, in the overdoped regime the high-energy background moves closer to the Fermi level, causing the main peak to appear slightly blue shifted and the shoulder to lose intensity.

To clarify the origin of the 70 meV peak and the 40 meV shoulder, in Fig. 3 we plot the squared electron–phonon matrix elements $|g(\mathbf{k},\mathbf{k}')|^2$, each of which represents the transition probability from a initial nodal state (corresponding to the combination of the Cu $d_{x^2-y^2}$ orbitals and the $p_\sigma$ orbitals of the planar O atoms) with wavevector $\mathbf{k}$ on the Fermi surface to a final state with wavevector $\mathbf{k}'$, associated with a given phonon with wavevector $\mathbf{q} = \mathbf{k}' - \mathbf{k}$. Our calculations show that the matrix elements corresponding to the half-breathing and the full-breathing modes are by far the largest among all vibrational modes (with the exception of the apical oxygen stretching mode; see Fig. 3 legend), thereby explaining the main peak in the self-energy at 70 meV. On the other hand, the 40 meV buckling/stretching modes show only moderately large matrix elements in comparison with the 70 meV branches. However, these modes connect the nodal region to a large portion of the Fermi surface, and therefore their contribution to the self-energy is enhanced by a significant phase space effect.

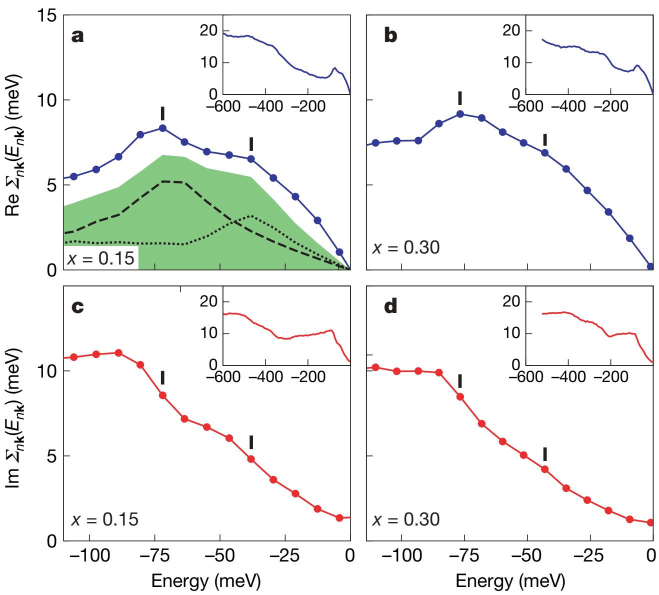



**Figure 2 | Calculated electron self-energy in LSCO due to the electron–phonon interaction. a, b,** Real parts of the self-energy for optimally doped ($x = 0.15$) and heavily overdoped ($x = 0.30$) LSCO, respectively, at 20 K. **c, d,** Imaginary parts of the self-energy for optimally and heavily overdoped LSCO, respectively. The inset axes in each panel are the same as those in the main plot. In **a**–**d** each inset shows the relevant (real or imaginary) part of the self-energy on an extended energy scale. We have computed the self-energy for electrons with parallel momentum **k** along the cut (0, 0, $k_\perp$)–($\pi$, $\pi$, $k_\perp$) (see Supplementary Fig. 1b) for three different values (0, $\pi a/c$, and $2\pi a/c$) of the normal component $k_\perp$. The variation of the self-energy along the $c$ axis is negligible along the cuts studied; therefore, we only show the results for the cut through the zone centre. The black markers in **a**–**d** indicate the main peak at ~70 meV and the shoulder at ~40 meV. In **a** the black dashed lines represent the contributions to the self-energy from the half-breathing and full-breathing Cu–O stretching modes, and the dotted lines represent the contributions to the self-energy from the breathing/stretching modes around 40 meV. The sum of these two contributions (green area) accounts for more than 80% of the real part of the self-energy. The imaginary part of the self-energy in the insets of **c** and **d** exhibits a step-like behaviour with a leading edge around 40–70 meV. This is a consequence of the Pauli exclusion principle: holes with energy below the threshold for phonon emission ~40–70 meV cannot make transitions to above the Fermi level by emitting a phonon, and thus exhibit longer lifetimes. Because the electronic density of states increases with increasing binding energy, a hole with large binding energy has an increased probability of decaying through phonon emission/absorption, owing to the larger number of final states available. The increased decay probability is reflected in the imaginary part of the self-energy and gives rise to the broad high-energy background in the real part through a Kramers–Kronig relation (see Supplementary Discussion and Supplementary Fig. 2).

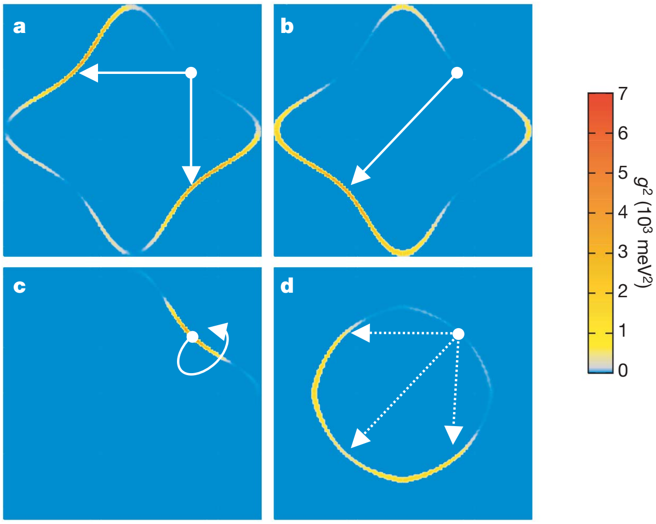



**Figure 3 | Electron–phonon matrix elements.** Intensity plots of the squared electron–phonon matrix elements $|g(\mathbf{k}, \mathbf{k}')|^2$, each of which represents the transition probability for the scattering from a state with wavevector **k** on the Fermi surface (indicated by the white dots) to a final state with wavevector **k**′ on the Fermi surface, for optimally doped LSCO. **a**, The matrix elements associated with the half-breathing phonons connect adjacent Fermi surface arcs with a maximum strength of $g^2 = 7.0 \times 10^3$ meV$^2$. **b**, The full-breathing modes connect Fermi arcs on the opposite sides of the Fermi surface with a maximum strength of $g^2 = 6.7 \times 10^3$ meV$^2$. **c**, The apical oxygen stretching mode also shows large matrix elements, of value $g^2 = 6.5 \times 10^3$ meV$^2$. However, this phonon gives rise only to momentum-conserving transitions (that is, with $\mathbf{k} = \mathbf{k}'$); hence, the number of allowed final states is negligible and its coupling to nodal holes is frustrated. **d**, The squared matrix element for the mixed buckling/stretching modes at a binding energy of 40 meV. These modes involve significant momentum transfer along the $c$ axis ($k'_\perp - k_\perp = 2\pi a/c$) and connect the nodal region to a large portion of the Fermi surface. The largest matrix elements are found at the intersection of the Fermi surface with the ($k_x, k_y, \pm 2\pi a/c$) planes. The largest matrix element in this case is $g^2 = 4.3 \times 10^3$ meV$^2$. The squared matrix elements in heavily overdoped LSCO are similar to those obtained at optimal doping (within 20%).

We now proceed to a quantitative comparison between our calculations and the photoemission data of ref. 1. The measured peaks provide the quasi-particle energy $E_{n\mathbf{k}}$ (referred to the Fermi energy) for a given parallel momentum **k**. The real part of the self-energy can be extracted from $\mathrm{Re}\,\Sigma_{n\mathbf{k}}(E_{n\mathbf{k}}) = E_{n\mathbf{k}} - \varepsilon_{n\mathbf{k}}$, where $\varepsilon_{n\mathbf{k}}$ denotes the non-interacting energy[18]. Along the nodal direction, the non-interacting dispersions are linear and we can assume, as is commonly done, that $\varepsilon_{n\mathbf{k}} = \hbar\mathbf{v}\cdot\mathbf{k}$, where the bare velocity **v** is to be determined. To extract the experimental self-energy we determined the bare velocity by fitting the raw data of figure 1a and d of ref. 1 to a simple Holstein self-energy model (see Supplementary Information). Using the experimentally determined **v**, we extracted the real part of the self-energy, which is shown in Fig. 4. A visual comparison between the theoretical self-energy and the experimental data indicates that the effect predicted by theory is considerably weaker. We note that this result does not depend on the choice of the bare band dispersions. When comparing the raw data of ref. 1 with our calculated electron dispersions fully renormalized by the electron–phonon interaction ($E_{n\mathbf{k}} = \varepsilon_{n\mathbf{k}} + \mathrm{Re}\,\Sigma_{n\mathbf{k}}(E_{n\mathbf{k}})$; see Fig. 4b, c), the discrepancy between experiment and theory is clearly visible.

To be quantitative and unambiguous, we extracted the electron–phonon coupling strength $\lambda$ by measuring the gradients of both the theoretical and the experimental self-energy data sets at low energy (in the limit of small binding energy the self-energy arising from the electron–phonon interaction is $\mathrm{Re}\,\Sigma_{n\mathbf{k}}(E_{n\mathbf{k}}) \approx -\lambda E_{n\mathbf{k}}$). We took the gradient at the Fermi level as an upper bound on $\lambda$, and the gradient between the Fermi level and a binding energy of 50 meV as a lower bound on $\lambda$. This procedure yielded the experimentally determined average electron–phonon couplings $\lambda_{\mathrm{expt}} = 1.00$–$1.32$ for the optimally doped sample at 20 K and $\lambda_{\mathrm{expt}} = 0.75$–$0.99$ for the overdoped sample at 20 K, consistent with the estimates provided in ref. 1. From the calculated self-energy, however, our analysis yielded the theoretical results $\lambda_{\mathrm{th}} = 0.14$–$0.22$ at optimal doping and $\lambda_{\mathrm{th}} = 0.14$–$0.20$ in the overdoped regime. Hence, the calculated renormalization of the nodal Fermi velocity is $\lambda_{\mathrm{expt}}/\lambda_{\mathrm{th}} = 5$–$7$ times smaller than the experimentally determined value.

Although there are numerous effects beyond the density-functional/Migdal formalism that ought to be carefully considered,

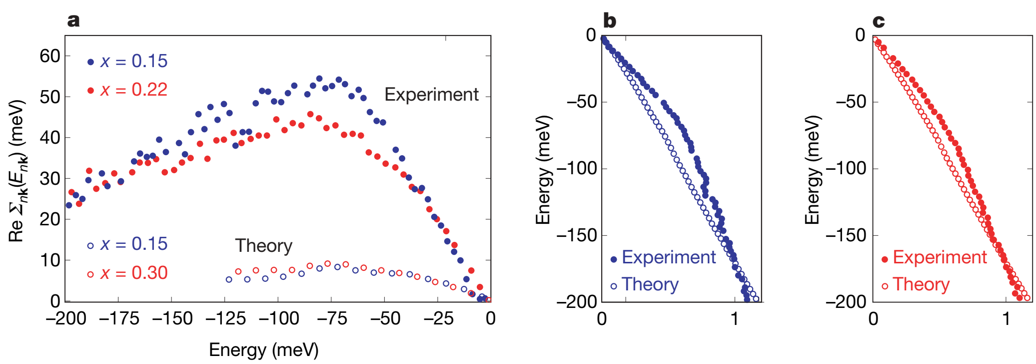



**Figure 4 | Comparison between theory and experiment.** **a**, Comparison between the real part of the electron self-energy, as obtained by angle-resolved photoemission spectroscopy experiments[1], and the first-principles theoretical self-energy arising from the electron–phonon interaction. From the raw data of Fig. 1a and d of ref. 1 we extracted the self-energy, assuming a linear dispersion and using a bare velocity of $\hbar|\mathbf{v}| = 3.4$ eV Å obtained by a fit to the data using a Holstein electron–boson self-energy model (see Supplementary Discussion and Supplementary Fig. 3; the precise value of the bare velocity is not critical to our conclusions). We show the experimental data at 20 K for optimally doped ($x = 0.15$) and overdoped ($x = 0.22$) LSCO, as well as the calculated self-energy in the optimally doped ($x = 0.15$) and heavily overdoped ($x = 0.30$) cases at the same temperature. **b**, **c**, Comparison between the quasi particle dispersion relations obtained from the peaks of the photoemission spectra and our calculations, including the renormalization due to the electron–phonon interaction (optimally doped (**b**), overdoped (**c**)). Following ref. 1, we normalized the photoelectron wavevector to one at the binding energy of 170 meV. The magnitude of the calculated self-energy is so small that the fine structures corresponding to the 40 meV and 70 meV excitations shown in Fig. 2 can hardly be seen. The Migdal approximation to the Feynman–Dyson perturbation theory (equation (1)) using the density functional theory band structure is reasonable for states along the nodal direction in the normal, superconducting and pseudogap phases, because the phonons which are responsible for the 70 meV peak in Fig. 2a connect the nodal regions mainly among themselves (Fig. 3a, b); therefore, the important transitions contributing to the self-energy are those between the Fermi surface arcs, where the gap function is very small.

the large discrepancy between theory and experiment stands out as a qualitative effect. Therefore, it is unlikely that such a discrepancy will be resolved by removing some of the approximations we adopted in our calculations (see Supplementary Information). Importantly, this disagreement is also found in the overdoped regime, where the effects of electron correlations are reduced. These observations lead us to conclude that the electron–phonon interaction in bulk LSCO cannot account for the magnitude of the experimentally observed kink.

Our results are consistent with previous first-principles calculations[4,5] on the electron–phonon coupling in the copper oxides. Brillouin zone-averaged coupling strengths $\lambda_{\mathrm{ave}} \approx 0.4$ and $\lambda_{\mathrm{ave}} = 0.27$ have been obtained using linear-response techniques for $Ca_{0.27}Sr_{0.63}CuO_2$ and $YBa_2Cu_3O_7$, respectively[4,5]. When we evaluate the same parameter using the standard definition, which corresponds to the first reciprocal energy moment of the Eliashberg function[19] $\alpha^2F(\omega)$, we obtain $\lambda_{\mathrm{ave}} = 0.4$ for optimally doped LSCO, in agreement with those studies. In addition, our results are fully consistent with very recent theoretical work[3] on the related compound $YBa_2Cu_3O_7$, where the renormalization of the nodal Fermi velocity was calculated to be 3–5 times smaller than the experimentally determined value.

The analysis of alternative descriptions involving spin excitations[26], as well as other models in which electron–phonon coupling plays a minor role in angle-resolved photoemission spectroscopy[27], is beyond the scope of the present work. However, the contribution of non-phononic mechanisms to the kink remains an important open question. We stress that the present study does not rule out theories of high-temperature superconductivity based on novel phonon mechanisms and topological approaches, such as the dopant percolative model[28].

Recently, a photoemission study of $Bi_2Sr_2CuO_6$ suggested that the apparent magnitude of the kink is enhanced by the curvature of the bare electronic bands[29], and that the electron–phonon interaction accounts for only a fraction of the kink. If we were to assume that effects not included in our theoretical treatment could moderately enhance our calculated self-energy, then the experimental and theoretical results on the phonon contribution could become consistent, in line with the recent observation of a conventional isotope effect (Z. X. Shen, personal communication). In any event, our calculations indicate an electron–phonon coupling considerably weaker than was previously suggested[1].

**Received 10 October 2007; accepted 21 February 2008.**


1. Lanzara, A. *et al.* Evidence for ubiquitous strong electron–phonon coupling in high-temperature superconductors. *Nature* **412,** 510–514 (2001).
2. Johnson, P.-D. *et al.* Doping and temperature dependence of the mass enhancement observed in the cuprate $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **87,** 177007 (2001).
3. Heid, R., Bohnen, K.-P., Zeyher, R. & Manske, D. Momentum dependence of the electron-phonon coupling and self-energy effects in $YBa_2Cu_3O_7$ within the local density approximation. *Phys. Rev. Lett.* (in the press); preprint at ⟨http://arxiv.org/abs/0707.4429v2⟩.
4. Savrasov, S.-Y. & Andersen, O.-K. Linear-response calculation of the electron-phonon coupling in doped $CaCuO_2$. *Phys. Rev. Lett.* **77,** 4430–4433 (1996).
5. Bohnen, K.-P., Heid, R. & Krauss, M. Phonon dispersion and electron-phonon interaction for $YBa_2Cu_3O_7$ from first-principles calculations. *Europhys. Lett.* **64,** 104–110 (2003).
6. Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75,** 473–541 (2003).
7. Sahrakorpi, S., Lindroos, M., Markiewicz, R.-S. & Bansil, A. Evolution of midgap states and residual three dimensionality in $La_{2-x}Sr_xCuO_4$. *Phys. Rev. Lett.* **95,** 157601 (2005).
8. Zhou, X.-J. *et al.* Dichotomy between nodal and antinodal quasiparticles in underdoped $(La_{2-x}Sr_x)CuO_4$ superconductors. *Phys. Rev. Lett.* **92,** 187001 (2004).
9. Perdew, J.-P. & Wang, Y. Pair-distribution function and its coupling-constant average for the spin-polarized electron gas. *Phys. Rev. B* **46,** 12947–12954 (1992).
10. Ihm, J., Zunger, A. & Cohen, M.-L. Momentum-space formalism for the total energy of solids. *J. Phys. C* **12,** 4409–4422 (1979).
11. Ino, A. *et al.* Doping-dependent evolution of the electronic structure of $La_{2-x}Sr_xCuO_4$ in the superconducting and metallic phases. *Phys. Rev. B* **65,** 094504 (2002).
12. Pickett, W.-E. Electronic structure of the high-temperature oxide superconductors. *Rev. Mod. Phys.* **61,** 433–512 (1989).
13. Baroni, S., De Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73,** 515–562 (2001).
14. McQueeney, R.-J. *et al.* Anomalous dispersion of LO phonons in $La_{1.85}Sr_{0.15}CuO_4$ at low temperatures. *Phys. Rev. Lett.* **82,** 628–631 (1999).
15. McQueeney, R.-J., Sarrao, J.-L., Pagliuso, P.-G., Stephens, P.-W. & Osborn, R. Mixed lattice and electronic states in high-temperature superconductors. *Phys. Rev. Lett.* **87,** 077001 (2001).
16. Pintschovius, L. *et al.* Lattice dynamical studies of HTSC materials. *Physica C* **185,** 156–161 (1991).
17. Falter, C. Phonons, electronic charge response and electron-phonon interaction in the high-temperature superconductors. *Phys. Status Solidi B* **242,** 78–117 (2005).
18. Hedin, L. & Lundqvist, S. in *Solid State Physics* Vol. 15 (eds Seitz, F., Turnbull, F. & Ehrenreich, H.) 1–181 (Academic, New York, 1969).
19. Grimvall, G. *The Electron-Phonon Interaction in Metals* Chs 3 5 (North-Holland, New York, 1981).
20. Allen, P.-B. & Mitrovich, B. *in Solid State Physics* Vol. 37 (eds Seitz, F., Turnbull, F. & Ehrenreich, H.) 1–92 (Academic, New York, 1982).
21. Valla, T., Fedorov, A.-V., Lee, J., Davis, J.-C. & Gu, J.-D. The ground state of the pseudogap in cuprate superconductors. *Science* **314,** 1914–1916 (2006).
22. Chen, X.-K., Irwin, J.-C., Trodahl, H.-J., Kimura, T. & Kishio, K. Investigation of the superconducting gap in $La_{2-x}Sr_xCuO_4$ by Raman spectroscopy. *Phys. Rev. Lett.* **73,** 3290–3293 (1994).
23. Marzari, N. & Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **56,** 12847–12865 (1997).
24. Giustino, F., Yates, J.-R., Souza, I., Cohen, M.-L. & Louie, S.-G. Electron-phonon interaction via electronic and lattice Wannier functions: superconductivity in boron-doped diamond reexamined. *Phys. Rev. Lett.* **98,** 047005 (2007).
25. Giustino, F., Cohen, M.-L. & Louie, S.-G. Electron-phonon interaction using Wannier functions. *Phys. Rev. B* **76,** 165108 (2007).
26. Eschrig, M. & Norman, M.-R. Neutron resonance: modeling photoemission and tunneling data in the superconducting state of $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **85,** 3261–3264 (2000).
27. Farid, B. On the break in the single-particle energy dispersion and the universal nodal Fermi velocity in the high-temperature copper oxide superconductors. *Phil. Mag.* **84,** 909–955 (2004).
28. Phillips, J. C. Self-organized networks and lattice effects in high-temperature superconductors. *Phys. Rev. B* **75,** 214503 (2007).
29. Meevasana, W. *et al.* Doping dependence of the coupling of electrons to bosonic modes in the single-layer high-temperature $Bi_2Sr_2CuO_6$ superconductor. *Phys. Rev. Lett.* **96,** 157003 (2006).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** The authors thank Z. X. Shen, A. Lanzara, Y.-W. Son, and C.-H. Park for discussions. This work was supported by the National Science Foundation and by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, of the US Department of Energy. Computational resources were provided by NPACI and NERSC. Part of the calculations were performed using modified versions of the packages Wannier90 and Quantum-ESPRESSO. The Fermi surfaces were rendered using the program XCrySDen (version 1.4.1).

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.G.L. (sglouie@berkeley.edu).

# LETTERS

# Eocene/Oligocene ocean de-acidification linked to Antarctic glaciation by sea-level fall

Agostino Merico[1,2], Toby Tyrrell[1] & Paul A. Wilson[1]

**One of the most dramatic perturbations to the Earth system during the past 100 million years was the rapid onset of Antarctic glaciation near the Eocene/Oligocene epoch boundary[1–3] (~34 million years ago). This climate transition was accompanied[3] by a deepening of the calcite compensation depth—the ocean depth at which the rate of calcium carbonate input from surface waters equals the rate of dissolution. Changes in the global carbon cycle[4], rather than changes in continental configuration[5], have recently been proposed as the most likely root cause of Antarctic glaciation, but the mechanism linking glaciation to the deepening of calcite compensation depth remains unclear. Here we use a global biogeochemical box model to test competing hypotheses put forward to explain the Eocene/Oligocene transition. We find that, of the candidate hypotheses, only shelf to deep sea carbonate partitioning is capable of explaining the observed changes in both carbon isotope composition and calcium carbonate accumulation at the sea floor. In our simulations, glacioeustatic sea-level fall associated with the growth of Antarctic ice sheets permanently reduces global calcium carbonate accumulation on the continental shelves, leading to an increase in pelagic burial via permanent deepening of the calcite compensation depth. At the same time, fresh limestones are exposed to erosion, thus temporarily increasing global river inputs of dissolved carbonate and increasing seawater $\delta^{13}$C. Our work sheds new light on the mechanisms linking glaciation and ocean acidity change across arguably the most important climate transition of the Cenozoic era.**

Evidence for Eocene ice rafting is compelling[6] but, despite speculation to the contrary[7], high-resolution oxygen isotope records[3,8] suggest that Cenozoic ice sheets approaching their modern size were not initiated on Antarctica until the Eocene/Oligocene boundary. Explanations for Antarctic glaciation at the Eocene/Oligocene boundary fall into two main categories: those invoking changes in ocean circulation through tectonic opening of Southern Ocean gateways, and those invoking changes in the global carbon cycle, with recent studies supporting the latter[4,9]. Proxy records of atmospheric $CO_2$ over the past 50 million years (Myr)[10] suggest an overall shift from high levels during the Eocene (~1,000 to ~2,000 parts per million by volume, p.p.m.v.) to lower levels thereafter. Support for a fall in $CO_2$ levels at the Eocene/Oligocene boundary comes from recent Ocean Drilling Program records[3]. These data show that deepening of the calcite compensation depth (CCD) across the Eocene/Oligocene boundary was rapid and took place in two jumps, in lockstep with the growth of Antarctic ice-sheets and carbon cycle perturbation—as tracked by changes in benthic foraminiferal calcite $\delta^{18}$O and $\delta^{13}$C, respectively. Yet the mechanism teleconnecting the onset of major Antarctic glaciation and CCD deepening is a subject of debate. At least four hypotheses have been invoked to link $\delta^{18}$O increase (glaciation), $\delta^{13}$C increase and CCD deepening (carbon cycle perturbation) across this key interval. These are H1, an increase in organic carbon burial rates[2,11–13]; H2, an increase in weathering of silicate rocks[13–15]; H3, an increase in global siliceous (versus calcareous) plankton export production[3]; and H4, a shift of global $CaCO_3$ sedimentation from shelf to deep ocean basins[3,16,17].

To test the power of these hypotheses to explain the Eocene/Oligocene transition, we have developed a biogeochemical box model of the global ocean (Fig. 1) with long-term control on the carbon cycle given by weathering processes[18]. A full suite of time series are output, including nutrient and phytoplankton concentrations, alkalinity, dissolved inorganic carbon (DIC), pH, $[CO_3^{2-}]$ for sea water, CCD, $\delta^{13}$C of burial fluxes plus atmospheric $CO_2$. In Fig. 2 we present the parameters best compared to the high-resolution data sets available: simulated CCD and $\delta^{13}$C in benthic foraminiferal calcite.

In some ways H1 represents the obverse scenario to present-day fossil fuel emissions—a period of rapid burial of organic carbon (carbon sequestration) leading to $CO_2$ extraction from the atmosphere–ocean system. Eocene/Oligocene boundary sections in the Southern Ocean[11,19] show increased accumulation rates of opal and benthic foraminifera suggesting that elevated marine organic carbon burial rates contributed to ice-sheet growth via acceleration of the biological pump and $CO_2$ removal[2,11]. Higher pH associated with $CO_2$ drawdown would have increased $[CO_3^{2-}]$ and deepened the CCD.

At least three different processes have been proposed to elevate organic carbon burial at the Eocene/Oligocene boundary. Arguably the simplest (H1a) is increased nutrient supply via more vigorous stirring in a cooler climate with a steeper meridional temperature gradient and a large Antarctic ice sheet[5,13,20]. We test H1a by a permanent 3.5-fold increase in vertical mixing rates in the model (mixing between surface and middle boxes, and between middle and deep boxes, KSM and KMD, respectively, in Fig. 1). The simulation fails (Fig. 2a and b, red lines) to reproduce the proxy records by predicting a permanent change in benthic $\delta^{13}$C and only a temporary change in the CCD (opposite to the observations). The latter reflects rapid depletion of the finite ocean nutrient inventory. The increase in ventilation also produces a pronounced temporary increase (by about 400 p.p.m.v.) in atmospheric $p_{CO_2}$ (Supplementary Fig. 1)—at odds with the onset of Antarctic glaciation.

A second process (H1b) is increased efficiency of organic carbon burial at the sea floor[2], perhaps attributable to slower bacterial activity in a cooler climate[12]. We test H1b by employing a permanent 50% increase in the fraction of surface-produced organic carbon (and phosphorus) that survives to be buried. The simulation fails (Fig. 2a and b, green lines) by predicting a permanent change in $\delta^{13}$C, whereas observations show a recovery. Using a temporary version of the same perturbation fails by predicting a temporary change in both $\delta^{13}$C and CCD.

[1]National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, UK. [2]GKSS-Forschungszentrum, Institute for Coastal Research, Max Planck Straße 1, 21502, Geesthacht, Germany.

We test a third process (H1c) for increased organic carbon burial (increased nutrient delivery to the oceans) by supplying an additional 20% of both phosphorus and silicate down rivers. In these runs we decouple the effect of this perturbation from the weathering feedback by delivering the extra silicate through some (unspecified) process that does not simultaneously deliver alkalinity to the ocean. These simulations produce a modest increase in $\delta^{13}C$ followed by a persistent decline, and a CCD change that is of the wrong sign (Fig. 2a and b, black lines). CCD shoaling in these runs is traceable to calcifiers that are also being stimulated by the influx of extra nutrients, hence leading to a mismatch between carbonate burial and delivery down rivers.

The second hypothesis (H2) tested is that early Oligocene glaciation cycles enhanced silicate weathering[13–15], reducing atmospheric $CO_2$. Weathering of silicate rocks is a sink for atmospheric $CO_2$ and a source of alkalinity to the oceans and so, potentially, might explain both climate cooling and CCD deepening. A natural additional consequence of increased silicate weathering on the continents is elevated riverine delivery of silicate leading to increased oceanic diatom production (see H3). However, to isolate the direct impact of silicate weathering on the carbon cycle, under H2, we allow silicate weathering to affect ocean alkalinity but not silicate levels. Under this hypothesis, a strong decrease in carbonate weathering, driven by climate cooling, causes $[CO_3^{2-}]$ to fall sharply, with a consequent shallowing of the CCD (Fig. 2d). Sensitivity analyses using different injections of alkalinity, and different dependences of the two weathering processes on $p_{CO_2}$ (Supplementary Fig. 6) produced the same conclusion (rejection of H2).

The increase in opal accumulation rates across the Eocene/Oligocene transition in the Southern Ocean[11,19] has also led to a third hypothesis (H3) in which the global balance shifted away from calcifiers towards silicifiers[3]. A plausible consequence of increased diatom dominance of phytoplankton assemblages is decreased abundance of other phytoplankton, including coccolithophores, because total production is limited by phosphate[21]. Assuming constant phosphorus and carbon river inputs, decreased surface production of $CaCO_3$ will eventually lead to an increase in $[CO_3^{2-}]$ and to a deepening of the CCD. In contrast to H2, we test this hypothesis by applying a permanent increase in the silicate riverine input to favour silicifiers against calcifiers. To decouple this perturbation from the weathering feedback, the extra silicate is delivered to the ocean through some (unspecified) process that does not simultaneously deliver alkalinity. The model captures some basic structure seen in the proxy records (for example, a permanent deepening of the CCD), but $\delta^{13}C$ does not recover to pre-excursion levels (Fig. 2e and f, red lines) when the perturbation is applied permanently. Applying the mechanism as a temporary perturbation results in a different mismatch with the data (simulated CCD shows only a temporary shift). An additional simulation in which increased silicate weathering is allowed to affect both alkalinity and silicate simultaneously (H2 in conjunction with H3) also produces unrealistic results with the effects of H2 outweighing those of H3.

A further hypothesis (H4) is a shift in the balance of global $CaCO_3$ burial from shelf to deep sea[3,16,17]. This hypothesis has two components[3]: glacioeustatic sea level fall[22] (by ~70 m) both reduces global shelf area and hence shelf burial of $CaCO_3$, and also exposes previously submarine shelf $CaCO_3$ to weathering. Modelling of the first component consists of a permanent decrease in shallow $CaCO_3$ burial. The second component, which has a temporary effect, consists of eroding and discharging to the ocean the newly exposed relatively fast-weathering[23] limestones. This extra discharge (~0.04–0.25 × $10^6$ Gt $CaCO_3$-C in total; peak annual flux increased by ~50% to 300% over baseline, depending on the $\delta^{13}C$ of this carbonate) compares to an estimated[23,24] ~20% to 50% increase in global carbonate weathering during the Last Glacial Maximum (LGM) over today. Intense $CaCO_3$ erosion at the Eocene/Oligocene boundary is

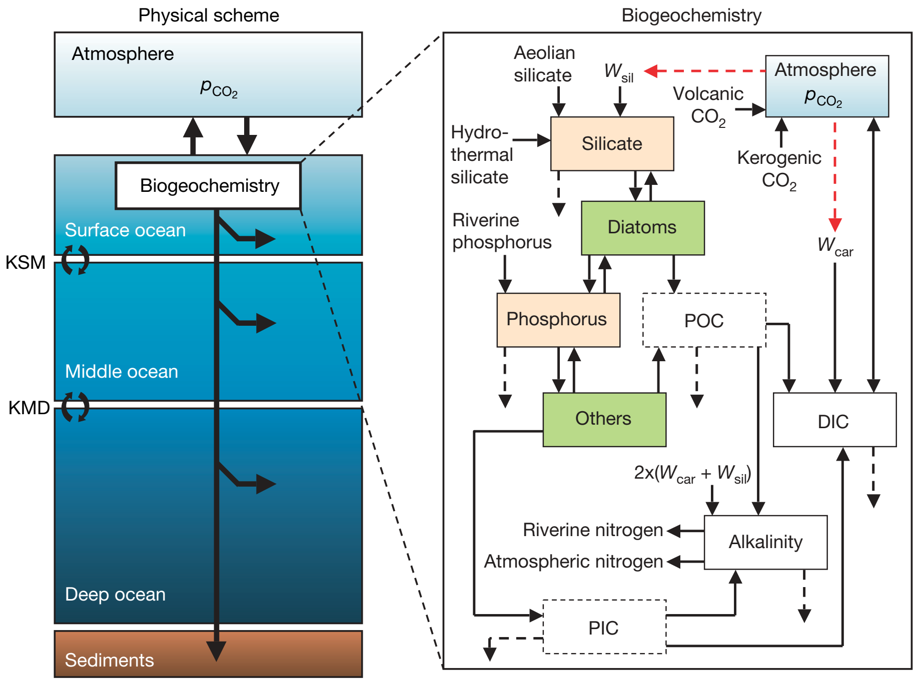



**Figure 1 | Our global biogeochemical box model.** The model comprises a three-box ocean module with an implicit sediment layer and an atmospheric box. The biology is modelled explicitly and includes diatoms and others (some of which represents calcifying phytoplankton), phosphorus and silicate. The distribution of DIC within the ocean is governed by physical, chemical and biological processes: the exchange of $CO_2$ between the atmosphere and the ocean, vertical advection of water, the organic carbon and calcium carbonate pumps and chemical speciation. Two of the processes reducing alkalinity are deposition of atmospheric nitrogen, and input of riverine nitrogen. In the biogeochemistry scheme, the dashed black arrows represent export out from the surface ocean. The silicate and carbonate feedbacks are indicated with dashed red arrows. PIC, particulate inorganic carbon; POC, particulate organic carbon; $W_{car}$, carbonate weathering; $W_{sil}$, silicate weathering; KSM and KMD, mixing between surface and middle boxes and between middle and deep boxes, respectively. The arrows in the oceanic boxes and sediments represent the various remineralization and sedimentation fluxes. A full list of variables and parameter values is given in Supplementary Tables 1 and 2.

based on a threefold greater Eocene inventory of shelf carbonate than today[17], and its likely much greater average solubility—having not undergone mineralogical stabilization during repeated precursor large-amplitude glacial lowstands. We model the Eocene/Oligocene input as a two-step function (Supplementary Fig. 3), in line with interpretations[3,4] of how the Antarctic ice sheet grew, but our findings are not sensitive to this implementation. In our 'best run', the CCD shows an overall permanent deepening with a maximum initial response of about 2 km; and $\delta^{13}C$ shows a temporary increase in agreement with the data (Fig. 2g and h, red lines).

We ascribe more importance to the similarity in form of the observed and modelled records than to absolute values for two reasons. First, the model captures the global signal, whereas the records come from a single site (albeit from the largest ocean). Second, the model is conservative in that Eocene global shelf-to-basin $CaCO_3$ partitioning is set to 45:55, close to the estimated present-day ratio[25], whereas the early Cenozoic shelf $CaCO_3$ sink was proportionately larger[17]. Accordingly, we conclude that H4 is compatible with the observations. A permanent shift in the CCD, towards deeper values, is induced by the requirement of the ocean to maintain the input/output balance of carbonate by a shelf-to-deep-basin switch in deposition. A change in $\delta^{13}C$ of the right sign and approximate magnitude is brought about by the weathering of extra neritic $CaCO_3$, having[26] a higher isotopic composition than the baseline $CaCO_3$ weathering and burial fluxes, both of which have substantial pelagic components (Supplementary Fig. 3).

It has been suggested[27] that increased weathering to the ocean of calcium ions is the critical process, via its effect on saturation state and the CCD. We test this alternative by re-running H4, but with $[Ca^{2+}]$ as a dynamic variable rather than a constant. In this experiment all $CaCO_3$ fluxes affect $[Ca^{2+}]$ as well as DIC and alkalinity. Given the large difference in residence times (calcium ~1 million years; DIC ~100,000 years) and concentrations (calcium ~10.6 mmol; DIC ~2.3 mmol), the addition of calcium produces a negligible difference to the model results (the green lines in Fig. 2g and h almost directly overlie their red line counterparts). We therefore discount this alternative.

The $\delta^{13}C$ and $CaCO_3$ MAR records from ODP Site 1218 provide strong constraints on the biogeochemistry of Eocene/Oligocene

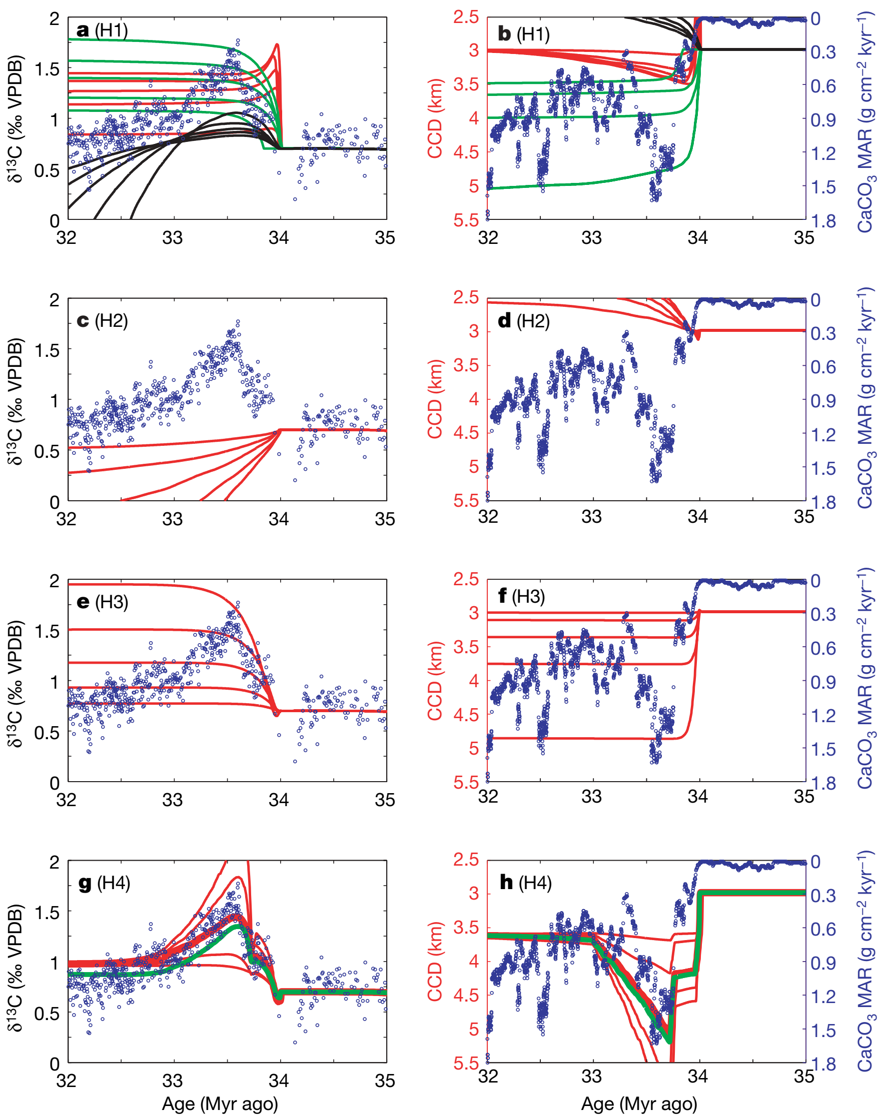



**Figure 2 | Model results simulating competing hypotheses and comparison to available data.** Carbon isotope ratios in benthic foraminifera (**a**, **c**, **e** and **g**) and CCD evolution (**b**, **d**, **f** and **h**) for families of runs simulating the four different hypotheses H1, H2, H3 and H4 compared to the data[3]. **a**, **b**, H1 shows increased organic carbon burial via: a more vigorous stirring of the ocean (red lines), increased efficiency of organic carbon burial at the sea floor (green lines), and increased nutrient delivery to the ocean (black lines). **c**, **d**, H2 shows enhanced weathering of silicate rocks (red lines). **e**, **f**, H3 shows silicifiers outcompeting calcifiers (red lines). **g**, **h**, H4 shows changes in shelf to deep basin $CaCO_3$ partitioning and erosion of newly exposed $CaCO_3$ as a result of sea-level fall, affecting the seawater saturation state via dumping of additional carbon to the ocean (red lines; we consider the thick red line to be our 'best run'), and via dumping of both carbon and calcium in a 1:1 ratio (green line). Extra calcium was added only to the 'best run'.

boundary events: a temporary increase in $\delta^{13}C$ followed by a full recovery, and a large deepening in the CCD, followed by only partial recovery such that a permanent deepening in the CCD of order 1 km is sustained. Any hypothesis must simultaneously account for both the permanent and the temporary changes. Only H4 is found to do this, when applied to our model.

Glacioeustatic sea-level fall associated with the growth of the Antarctic ice sheet exposes widespread limestones to erosion; this leads to a one-off 'dump' of dissolved inorganic carbon and alkalinity into the ocean, increasing $[CO_3^{2-}]$, deepening the CCD and increasing seawater $\delta^{13}C$. The reduction in shelf habitat for neritic benthic calcifiers leads to a large reduction in $CaCO_3$ burial state in shallow seas, with superimposed shorter term fluctuations as the East Antarctic ice sheet varies in size in response to orbital forcing[3,28]. Because river inputs of dissolved $CaCO_3$ must match global $CaCO_3$ burial[29], the decline in shallow-sea burial forces a compensating increase in deep-sea burial through a shift in state to a deeper CCD.

In nature, we might expect shelf to basin $CaCO_3$ fractionation to act in concert with other processes but additional runs to test H4 in combination with H1, H2 and H3 each resulted in poorer simulations of the high-resolution data sets available (Supplementary Fig. 2). Improved $p_{CO_2}$ records are needed. Our simulations show that, acting alone, H4 brings about a large but transient $p_{CO_2}$ decline (Supplementary Fig. 1). If CCD deepening at the Eocene/Oligocene boundary marks a permanent shift to a lower $p_{CO_2}$ regime then additional processes are indicated. Either way, our findings in support of a sea-level-led hypothesis (H4) lend weight to the view[3] that rapid carbon-cycle perturbation associated with CCD deepening at the Eocene/Oligocene boundary did not trigger Antarctic glaciation. Instead, rapid glaciation, itself driven by slow long-term $CO_2$ drawdown via silicate weathering[4] and orbital forcing[3] (inhibiting warm Antarctic summers), is capable of triggering a chain of carbon cycle responses, including CCD deepening, $\delta^{13}C$ perturbation and associated further (ice-sheet-stabilizing) rapid $p_{CO_2}$ drawdown, cooling and aridification[30].

## METHODS SUMMARY

Our biogeochemical box model of the global ocean includes two phytoplankton groups: diatoms and others (the latter including calcifiers) and the cycling of silicate, phosphorus, DIC and alkalinity (Fig. 1). The model simulates dynamic changes in the CCD and ocean carbon chemistry is linked, through air–sea gas exchange, with atmospheric $CO_2$. Both $^{12}C$ and $^{13}C$ are simulated in each reservoir and flux. Long-term control on the carbon cycle is given by weathering processes[18]. A full list of variables and parameter values is given in Supplementary Tables 1 and 2.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 21 September 2007; accepted 19 February 2008.**

1. Miller, K. G., Wright, J. D. & Fairbanks, R. G. Unlocking the icehouse: Oligocene-Miocene oxygen isotopes, eustasy, and margin erosion. *J. Geophys. Res.* **96**, 6829–6849 (1991).
2. Zachos, J. C., Quinn, T. M. & Salamy, K. A. High-resolution ($10^4$ years) deep-sea foraminiferal stable isotope records of the Eocene-Oligocene climate transition. *Palaeoceanography* **11**, 251–266 (1996).
3. Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).
4. DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica triggered by declining atmospheric $CO_2$. *Nature* **421**, 245–249 (2003).
5. Kennett, J. P. & Shackleton, N. J. Oxygen isotopic evidence for the development of the psychrosphere 38 Myr ago. *Nature* **260**, 513–515 (1976).
6. Eldrett, J. S., Harding, I. C., Wilson, P. A., Butler, E. & Roberts, A. P. Continental ice in Greenland during the Eocene and Oligocene. *Nature* **446**, 176–179 (2007).
7. Tripati, A., Backman, J., Elderfield, H. & Ferretti, P. Eocene bipolar glaciation associated with global carbon cycle changes. *Nature* **436**, 341–346 (2005).
8. Edgar, K. M., Wilson, P. A., Sexton, P. S. & Suganuma, Y. No extreme bipolar glaciation during the main Eocene calcite compensation shift. *Nature* **448**, 908–911 (2007).
9. Lyle, M., Gibbs, S., Moore, T. C. & Rea, D. K. Late Oligocene initiation of the Antarctic Circumpolar Current: Evidence from the South Pacific. *Geology* **35**, 691–694 (2007).
10. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–603 (2005).
11. Salamy, K. A. & Zachos, J. C. Latest Eocene-Early Oligocene climate change and Southern Ocean fertility: inferences from sediment accumulation and stable isotope data. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **145**, 61–77 (1999).
12. Olivarez Lyle, A. & Lyle, M. W. Missing organic carbon in Eocene marine sediments: is metabolism the biological feedback that maintains end-member climates? *Paleoceanography* **21**, PA2007, doi:10.1029/2005PA001230 (2006).
13. Zachos, J. C. & Kump, L. R. Carbon cycle feedbacks and the initiation of Antarctic glaciation in the earliest Oligocene. *Glob. Planet. Changes* **47**, 51–66 (2005).
14. Zachos, J. C., Opdyke, B. N., Quinn, T. M., Jones, C. E. & Hallid, A. N. Early cenozoic glaciation, Antarctic weathering, and seawater $^{87}Sr/^{86}Sr$: is there a link? *Chem. Geol.* **161**, 165–180 (1999).
15. Ravizza, G. E. & Peucker-Ehrenbrinck, F. The marine $^{187}Os/^{188}Os$ record of the Eocene-Oligocene transition: the interplay of weathering and glaciation. *Earth Planet. Sci. Lett.* **210**, 151–165 (2003).
16. Kump, L. R. & Arthur, M. A. in *Tectonic Uplift and Climate Change* (ed. Ruddiman, W. F.) 399–426 (Plenum, New York, 1997).
17. Opdyke, B. N. & Wilkinson, B. H. Surface area control of shallow cratonic to deep marine carbonate accumulation. *Paleoceanography* **3**, 685–703 (1988).
18. Walker, J. C. G. & Kasting, J. F. Effects of fuel and forest conservation on future levels of atmospheric carbon dioxide. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **97**, 151–189 (1992).
19. Diester-Haass, L. Middle Eocene to early Oligocene paleoceanography of the Antarctic Ocean (Maud Rise, ODP Leg 113, Site 689): change from a low to a high productivity ocean. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **113**, 311–334 (1995).
20. Kennett, J. P. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic Ocean, and their impact on global paleoclimate. *J. Geophys. Res.* **82**, 3843–3860 (1977).
21. Tyrrell, T. The relative influence of nitrogen and phosphorus on oceanic primary prduction. *Nature* **400**, 525–531 (1999).
22. Pekar, S. F., Christie-Blick, N., Kominz, M. A. & Miller, K. G. Calibration between eustatic estimates from backstripping and oxygen isotopic records for the Oligocene. *Geology* **30**, 903–906 (2002).
23. Gibbs, M. T. & Kump, L. R. Global chemical erosion during the last glacial maximum and the present: sensitivity to changes in lithology and hydrology. *Paleoceanography* **9**, 529–543 (1994).
24. Munhoven, G. Glacial-intergalacial changes of continental weathering: estimates of the related $CO_2$ and $HCO_3$- flux variations and their uncertainties. *Glob. Planet. Changes* **33**, 155–176 (2002).
25. Milliman, J. D. Production and accumulation of calcium carbonate in the ocean: budget of a nonsteady state. *Glob. Biogeochem. Cycles* **7**, 927–957 (1993).
26. Swart, P. K. & Eberli, G. The nature of $\delta^{13}C$ of periplatform sediments: Implications for stratigraphy and the global carbon cycle. *Sedim. Geol.* **175**, 115–129 (2005).
27. Rea, D. K. & Lyle, M. W. Paleogene calcite compensation depth in the eastern subtropical Pacific: answers and questions. *Paleoceanography* **20**, PA1012, doi:10.1029/2004PA001064 (2005).
28. Pälike, H. *et al.* The heartbeat of the Oligocene climate system. *Science* **314**, 1894–1898 (2006).
29. Zeebe, R. & Westbroek, P. A simple model for the CaCO3 saturation state of the ocean: The "Strangelove", the "Neritan", and the "Cretan" ocean. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2003GC000538 (2003).
30. Dupont-Nivet, G. *et al.* Tibetan plateau aridification linked to global cooling at the Eocene–Oligocene transition. *Nature* **445**, 635–638 (2007).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank P. Sexton for discussions. We gratefully acknowledge R. DeConto, J. Kasting, G. Munhoven, H. Pälike and J. Walker for their comments on various aspects of our model, K. Wirtz for support and the UK Natural Environment Research Council for funding. We also thank R. Zeebe for comments on the manuscript.

**Author Contributions** All three authors contributed equally to this work.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.M. (agostino.merico@gkss.de).

## METHODS

**Description of the model.** The ocean component, developed from previous work[18,31], includes three vertically stacked boxes: the surface (0–100 m), which represents the euphotic zone, a middle box (100–500 m), which represents the mixed surface layer above the annual thermocline, and a deep box (500–3,730 m), which represents the deep layer below the annual thermocline. The model represents an average water column down to the seabed, has a spatially and temporally averaged input of nutrients, DIC and alkalinity, and does not take into account any latitudinal or horizontal variations. The distribution of DIC within the ocean is governed by physical, chemical and biological processes: the exchange of $CO_2$ between the atmosphere and the ocean, riverine input of DIC, biological uptake of $CO_2$ into biomass, remineralization and burial of biomass, precipitation and dissolution of $CaCO_3$ and mixing processes between the three layers. The processes governing the distribution of alkalinity in the ocean are: riverine input of bicarbonate and nitrate, precipitation of $CaCO_3$ by organisms and the biological uptake of nitrate, the dissolution of $CaCO_3$ deeper in the ocean and the remineralization of nitrate, burial of $CaCO_3$ and mixing processes between the three layers. The production of $CaCO_3$ in the surface ocean is linked to the production of organic matter through the 'rain ratio', which is the molar ratio of particulate inorganic carbon (PIC) export from the surface layer to particulate organic carbon (POC) export.

The silicate and carbonate weathering processes are modelled (according to ref. 18) using: $W_{car} = f_{car} \times (p_{CO_2}/p_{CO_{2(ini)}})$ and $W_{sil} = f_{sil} \times (p_{CO_2}/p_{CO_{2(ini)}})^{0.3}$. $W_{car}$ and $W_{sil}$ are normalized to the initial $p_{CO_2}$ value: $p_{CO_2(ini)}$. Input terms to the atmosphere also include a weathering process, the oxidation of fossil POC from rocks (kerogenic input), as well as $CO_2$ from volcanic outgassing (volcanic input).

The model includes the calculation of the dynamic CCD and aragonite compensation depth (ACD). Ocean hypsometry was obtained from the ETOPO5 data set at $5 \times 5$ min resolution across the global ocean. The relative fractions of ocean area at different depths were calculated by summing over the data set. To prevent unrepresentative biasing towards polar latitudes, each grid cell was weighted by the cosine of its latitude. The critical carbonate ion concentration ($[CO_3^{2-}]_{crit}$) below which sea water is undersaturated is a function of depth and is dependent on the crystal structure, specifically, calcite (mainly foraminifera and coccolithophores) or aragonite (mainly neritic calcifiers). The model calculates deep carbonate ion concentration, and the following equations[32] (for calcite and aragonite respectively) $[CO_3^{2-}]_{crit}(z) = 88.7\exp[0.189(z - 3.82)]$ and $[CO_3^{2-}]_{crit}(z) = 117.5\exp[0.176(z - 3.06)]$ are used to calculate the depth at which $[CO_3^{2-}]_{crit} = [CO_3^{2-}]$. The ocean hypsometry is then used to calculate the fraction of ocean area above or below the CCD and the ACD, and hence the proportions of the sinking $CaCO_3$ flux that are buried or dissolved.

The rate of $CaCO_3$ dissolution in the model is calculated solely as the product of the $CaCO_3$ sinking flux and the fraction of sea floor receptive to $CaCO_3$ burial. Previous carbon cycle studies[33–36] have included chemical erosion and/or respiration-driven dissolution. However, although chemical erosion of previously deposited $CaCO_3$ in the bioturbated zone must be considered for an ocean becoming increasingly acidic, it is not crucial for this study, in which the CCD is deepening. Respiratory dissolution does not generally decouple the lysocline from the saturation horizon[34] and so was not included. Following ref. 29, dissolution of $CaCO_3$ is modelled without explicitly modelling sediment chemistry.

Stable isotope composition of the shell calcite of most species of foraminifera reveals offsets from calcite precipitated in equilibrium with the ambient bottom water[37]. These offsets have been attributed to so-called vital effects. Evidence from culture experiments[38] suggests a significant effect of seawater carbonate ion concentration on the stable isotope composition of planktic foraminifera. This aspect was also taken into account in our model by using[38]: $\delta^{13}C_{shell} = \delta^{13}C_{DIC} - 0.008([CO_3^{2-}] - 300)$. We also include the $CO_2$ effect on the $\delta^{13}C$ of POC by using the following[39]: $\delta^{13}C_{org} = \delta^{13}C_{DIC} - 9866/T_K + 24.12 - 17\log_{10}([CO_{2(aq)}]) + 3.4$.

We note that whereas the model is based on a rather simple physical scheme of the global ocean (three vertical boxes), a previous study[4] (using a complex general circulation model with coupled components for atmosphere, ocean, ice sheet and sediment) has shown that physical processes (driven for instance by changes in continental configuration) had only a secondary role in the Eocene/Oligocene climate transition, compared to the processes controlling $CO_2$ concentration. A sensitivity analysis investigation suggests that the results obtained here are insensitive to model simplifications and assumptions (Supplementary Information).

**Steady-state and implementation of perturbations.** The initial time of the simulations is the late Eocene (35 Myr ago). The model is run for 3 Myr across the Eocene/Oligocene boundary up to the early Oligocene (32 Myr ago). We start the model in the steady state, then apply a perturbation in accordance with each hypothesis at the Eocene/Oligocene boundary. As part of the sensitivity analysis, we generated a family of runs for each hypothesis by changing parameter values. We choose to establish the initial steady state at an atmospheric $p_{CO_2}$ of 1,000 p.p.m.v. ($p_{CO_2\,(ini)}$), in keeping with proxy records[10]. In accordance with the data available[3] from the Pacific ('global') ocean, this steady-state is characterized by $\delta^{13}C$ in benthic foraminifera of +0.7‰ VPDB (Vienna Pee-Dee Belemnite standard) and depths for the CCD and ACD of about 3 km and 0.7 km, respectively. Major fluxes resulting from this baseline steady state are reported in Supplementary Table 3.


31. Chuck, A., Tyrrell, T., Totterdell, I. J. & Holligan, P. M. The oceanic response to carbon emissions over the next century: investigation using three ocean carbon cycle models. *Tellus B* **57**, 70–86 (2005).
32. Jansen, H., Zeebe, R. E. & Wolf-Gladrow, D. A. Modeling the dissolution of settling $CaCO_3$ in the ocean. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2000GB001279 (2002).
33. Archer, D. A data-driven model of the calcite lysocline. *Glob. Biogeochem. Cycles* **10**, 511–526 (1996).
34. Sigman, D. M., McCorkle, D. C. & Martin, W. R. The calcite lysocline as a constraint on glacial/interglacial low-latitude production changes. *Glob. Biogeochem. Cycles* **12**, 409–427 (1998).
35. Broecker, W. S. & Peng, T.-H. The role of CaCO3 compensation in the glacial to interglacial atmospheric CO2 change. *Glob. Biogeochem. Cycles* **1**, 15–29 (1987).
36. Ridgwell, A. *et al.* Marine geochemical data assimilation in an efficient Earth System Model of global biogeochemical cycling. *Biogeosciences* **4**, 87–104 (2007).
37. Rohling, E. J. & Cooke, S. in *Modern Foraminifera* (ed. Sen Gupta, B. K.) 239–258 (Kluwer Academic, Dordrecht, 1999).
38. Spero, H. J., Bijma, J., Lea, D. W. & Bemis, B. E. Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes. *Nature* **390**, 497–500 (1997).
39. Hofmann, M., Broecker, W. S. & Lynch-Stieglitz, J. Influence of a $[CO_{2(aq)}]$ dependent biological C-isotope fractionation on glacial $^{13}C/^{12}C$ ratios in the ocean. *Glob. Biogeochem. Cycles* **13**, 873–883 (1999).

LETTERS

# Hydrous silicate melt at high pressure

Mainak Mookherjee[1], Lars Stixrude[2] & Bijaya Karki[3]

**The structure and physical properties of hydrous silicate melts and the solubility of water in melts over most of the pressure regime of Earth's mantle (up to 136 GPa) remain unknown. At low pressure (up to a few gigapascals) the solubility of water increases rapidly with increasing pressure[1], and water has a large influence on the solidus temperature, density[2], viscosity[3] and electrical conductivity. Here we report the results of first-principles molecular dynamics simulations of hydrous $MgSiO_3$ melt. These show that pressure has a profound influence on speciation of the water component, which changes from being dominated by hydroxyls and water molecules at low pressure[4] to extended structures at high pressure. We link this change in structure to our finding that the water–silicate system becomes increasingly ideal at high pressure: we find complete miscibility of water and silicate melt throughout almost the entire mantle pressure regime. On the basis of our results, we argue that a buoyantly stable melt at the base of the upper mantle would contain approximately 3 wt% water and have an electrical conductivity of 18 S m$^{-1}$, and should therefore be detectable by means of electromagnetic sounding.**

The hydrogen bond is not well described by the dominantly ionic, atomistic models that have most often been applied to the study of silicate melts. First-principles molecular dynamics simulations are more costly in terms of computer time but have the important advantage of making no a priori assumptions regarding the nature of the bonding or the shape of the charge density. Density functional theory, on which our simulations are based, has been successfully applied to the study of a number of hydrous crystalline silicates and oxides[5] and to the liquid $SiO_2$-$H_2O$ system at ambient pressure[6].

In our first-principles molecular dynamics simulations of a hydrous $MgSiO_3$-$H_2O$ liquid with 10 wt% water, the electronic structure, the Hellman–Feynman forces acting on the nuclei, and thermodynamic properties are computed at each time step in the local density and pseudopotential approximations. Atomic trajectories extracted from the simulations reveal the structure, speciation and dynamics of the melt. To quantify the influence of the water component on physical properties, we compare our results with a previous first-principles molecular dynamics study of anhydrous $MgSiO_3$ liquid[7].

Inspection of the equilibrated liquid structure shows that pressure has a large influence on the speciation of the water component (Fig. 1). Whereas at low pressure we find mostly hydroxyls and water molecules, at high pressure we see a much greater variety of species. These include Si–O–H–O–Si polyhedral linkages, –O–H–O–H– chains and O–H–O edge decoration of $SiO_6$ octahedra, the last of which was suggested in a previous experimental study of glasses[8]. We note another feature at low pressure that had not been anticipated experimentally: a strong preference for water molecules to bond to $Mg^{2+}$ cations. This chemical association, together with the decrease in the proportion of hydroxyls with increasing pressure, may explain the increase in the $MgO/SiO_2$ ratio of eutectic melts in the $MgO$-$SiO_2$-$H_2O$ system with increasing pressure up to 15 GPa (ref. 9).

The water component becomes more interconnected with increasing pressure, transforming from isolated molecular species at low pressure to a structure more reminiscent of bulk phases of water at higher pressure (Fig. 2a). The H–O coordination number varies in value from almost one at low pressure to two at the highest pressures investigated. This coordination number is similar to that in the high-pressure form of water[10] and in ice in which the structure is a complete three-dimensional network of nearly symmetric hydrogen bonds (ice X). Simultaneously, as compression increases, the very open silica framework gives way to a much more densely packed structure in which little free volume remains to accommodate highly polar molecular species: molecules are squeezed out as pressure increases, and are replaced by extended structures.

The addition of water disrupts the intermediate-range silicate structure: the O–Si coordination number of the hydrous melt is substantially less than that of the anhydrous melt, demonstrating that added water disrupts interpolyhedral linkages (Fig. 2b). Such disruption is consistent with experimental evidence at low pressure for the decrease in viscosity with increasing water content in silicate melts. Because we find water-induced disruption of the silicate network at all pressures, we anticipate that addition of water will decrease the viscosity of silicate melts at high pressure as well. In contrast, the Si–O coordination number in hydrous and anhydrous melts is identical at all pressures. Mg–O coordination numbers are similarly unaffected by the addition of water at pressures greater than 20 GPa; at lower pressures the addition of water decreases the Mg–O coordination number through the interaction of water molecules with Mg cations.

We find that protons are highly mobile, presenting a possible means of geophysical detection of deep melt. The diffusivity of hydrogen is an order of magnitude greater than that of the other ions. From our simulations, $D$, the self-diffusion coefficient of hydrogen, is adequately described by the Arrhenius relation (Fig. 3)

$$D = D_0 \exp\left(-\frac{E^* + PV^*}{kT}\right) \qquad (1)$$

where $E^*$ denotes the activation energy, $V^*$ denotes the activation volume, $T$ denotes the temperature, $P$ denotes the pressure, $D_0$ denotes the value of $D$ in the limit of infinite temperature and $k$ denotes the Boltzmann constant. The electrical conductivity $\sigma$ due to charged mobile species is given by the Nernst–Einstein relation

$$\sigma = \frac{Dxq^2}{kTH_R} \qquad (2)$$

where $x$ is the number of carriers per unit volume, the Haven ratio $H_R$ approaches a value of one for small values of $x$ (ref. 11) and $q$ is the electrical charge of the carrier.

Assuming that protons are the primary charge carriers and that $H_R = 1$, from equations (1) and (2) we estimate a conductivity of $59c/10$ S m$^{-1}$ under conditions thought to represent the mantle at a depth of 410 km (that is, a pressure of 14 GPa and a temperature of

[1]Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA. [2]Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK. [3]Department of Computer Science and Department of Geology and Geophysics, Louisiana State University, Baton Rouge, Louisiana 70803, USA.

1,800 K), where $c$ is the percentage mass fraction of water in the melt. This corresponds to approximately 18 S m$^{-1}$ for a neutrally buoyant melt, which at this depth has $c = 3$ wt% (see below). The electrical conductivity of hydrous silicate melt has not been measured at pressures greater than 1 GPa. Because the expected water concentration in the solid mantle on either side of a hydrous melt layer is much less than 1%, the melt layer would represent a large conductance anomaly (of 18,000 S, assuming that $\sigma = 18$ S m$^{-1}$ and there is a 20-km-thick partial melt layer with 5% neutrally buoyant hydrous melt[12]) that should be detectable by means of electromagnetic sounding. Indeed, recent inversions of electromagnetic data, although non-unique, fit a value of total conductance similar to the values we find[13].

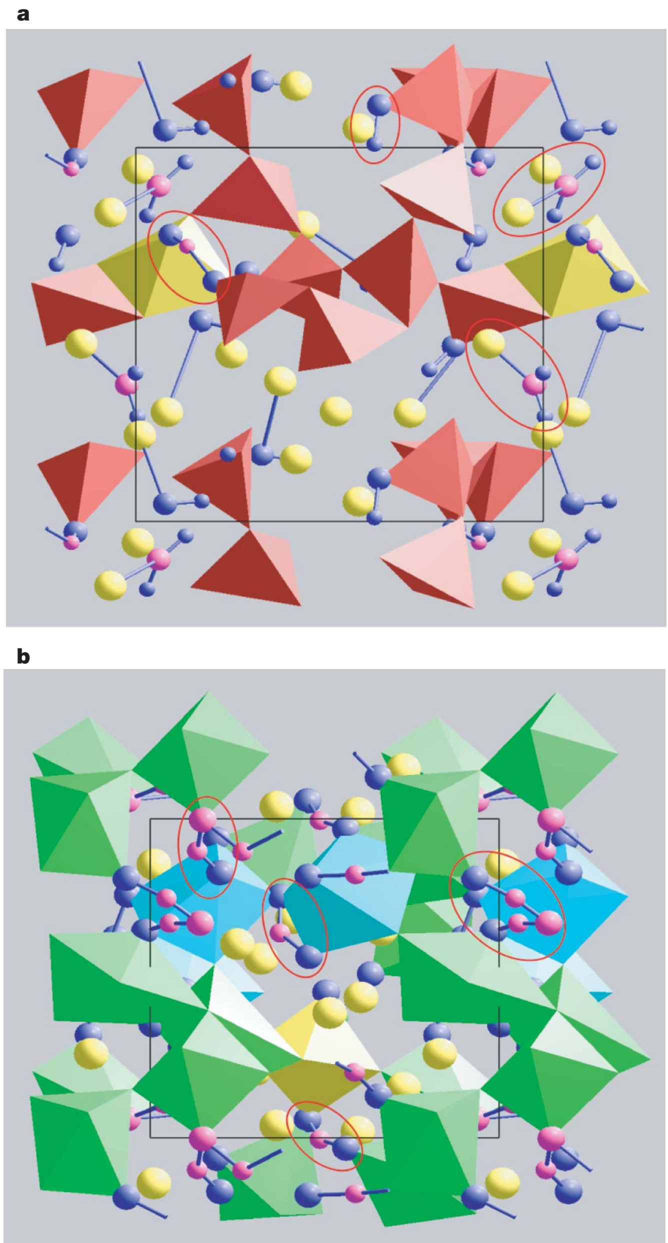



**Figure 1 | Structure and speciation of hydrous silicate melt.** **a**, 3,000 K, $V/V_x = 1$; **b**, 3,000 K, $V/V_x = 0.5$. Red, yellow, green and cyan polyhedra represent four-, five-, six- and sevenfold Si–O coordination environments, respectively. Yellow spheres represent Mg atoms. Other spheres represent H atoms (small spheres) and those O atoms (large spheres) not solely bonded to Si; they are coloured pink if they are bonded to at least two other H atoms or O atoms, and are coloured blue if they are bonded to only one other H or O atom. Red ellipses highlight species discussed in the text, including hydroxyls (O–H in **a**), water molecules bound to Mg ions (Mg–O–$H_2$ in **a**), interpolyhedral bridges (Si–O–H–O–Si in **b**), polyhedral edge decoration (O–H–O in **b**) and a Si–O–H–O–H–O–Si interpolyhedral bridge (in **b**). The apparent equilibrium constant of the reaction[4] $H_2O + O = 2OH^-$ at 3,000 K and near-ambient pressure computed from our simulations is 3.1, which is in close agreement with the value (2.3) extrapolated from experimental measurements on rhyolite compositions[27] at much lower temperatures and water contents, and is consistent with the experimentally observed tendency of water molecules increasingly to dissociate in the melt as temperature increases. The molecular dynamics snapshots were rendered using the visualization system of Bhattarai *et al.*[28]. In each panel, the black outline represents the primary simulation cell.

The equation of state of the hydrous liquid, like that of the anhydrous liquid, differs from that of most crystalline mantle phases in an important way: the thermal pressure increases markedly on compression (Fig. 4). The hydrous liquid equation of state can be described by the Mie–Grüneisen form

$$P(V,T) = P_c(V,T) + \frac{\gamma}{V} C_v (T - T_o)$$

where $P_c$ denotes the reference isotherm, taken to lie at $T_o = 3{,}000$ K, $\gamma$ denotes the Grüneisen parameter and $C_v$ denotes the isochoric heat capacity. These quantities are obtained from the simulation: as the internal energy $E$ and pressure $P$ vary linearly with temperature along isochores (to within our uncertainty), the values are calculated at each volume from $\gamma = V(\partial P/\partial E)_V$ and $C_v = (\partial E/\partial T)_V$. The isochoric heat capacity decreases by about 8% upon twofold compression, from $3.63 \pm 0.21$ $Nk$ to $3.34 \pm 0.11$ $Nk$, where $N$ is the number of atoms. The value of $\gamma$ increases by a factor of three over the same compression range. This behaviour is opposite to that of mantle

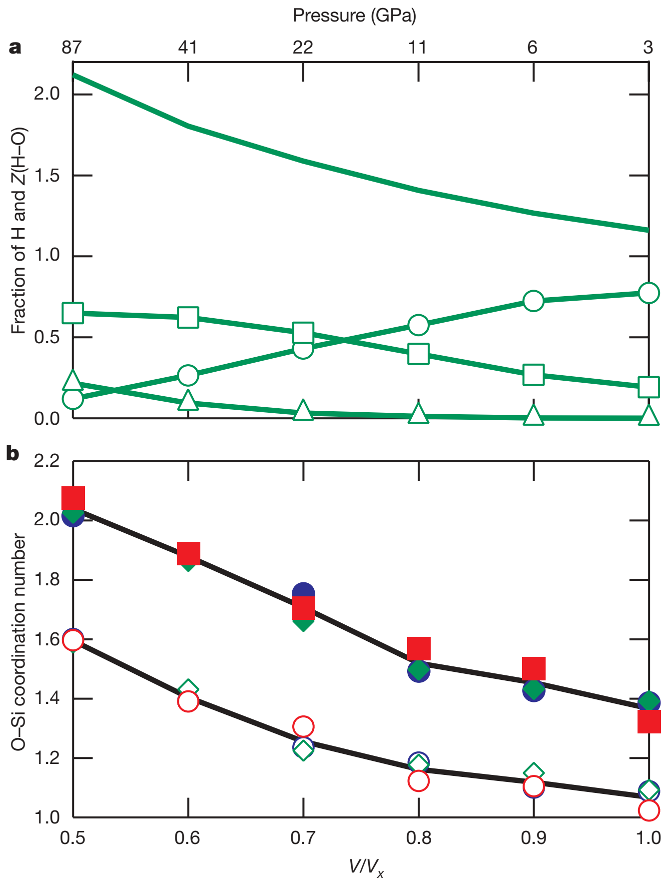



**Figure 2 | H–O and O–Si coordination numbers.** **a**, Fraction of protons bonded to one (open circles), two (squares) and three (triangles) O atoms, and the mean H–O coordination number $Z$(H–O) (solid line without symbols) at 4,000 K. The simulated pressure at 4,000 K is indicated along the top axis. **b**, Evolution of O–Si coordination number upon compression, for anhydrous[7] (filled symbols) and hydrous (open symbols) melts at 3,000 K (blue), 4,000 K (green) and 6,000 K (red). Black lines represent temperature-averaged values.

crystalline phases and similar to that of anhydrous $MgSiO_3$ liquid[7]: within our uncertainty, the values of $\gamma$ for hydrous and anhydrous melts are identical over the entire volume range explored.

The density of the hydrous melt is less than that of the anhydrous melt under all conditions explored here (Fig. 4a), but the density contrast is small enough that hydrous melts may be neutrally buoyant in the mantle, providing a possible explanation of a low-velocity layer found on top of the 410-km discontinuity[12]. The density contrast is remarkably constant across the pressure–temperature range of the mantle, varying from 0.3 to 0.4 g cm$^{-3}$ between 1 and 100 GPa and 3,000 and 6,000 K. We estimate the amount of water that would produce a neutrally buoyant melt at a depth of 410 km as $c/10=(\rho_{410}-\rho_{\mathrm{anhydrous}})/\Delta\rho_{H_2O}$, where $\rho_{410}=3.54$ g cm$^{-3}$ is the density at the base of Earth's upper mantle[14], $\rho_{\mathrm{anhydrous}}=3.66$ g cm$^{-3}$ is the density of a partial melt of anhydrous peridotite[15] at 13.4 GPa and 1,873 K, and $\Delta\rho_{H_2O}=-0.35$ g cm$^{-3}$ is the density contrast that we find for 10 wt% water. We find from this analysis that $c=3$ wt%, consistent with recent estimates[16,17].

The partial molar volume of water in the silicate melt, $\bar{V}_{H_2O}$, asymptotically approaches that of pure water, $V_{H_2O}$, as pressure increases (Fig. 4b). At the highest pressure that we studied, $\bar{V}_{H_2O}$ and $V_{H_2O}$ are identical to within our uncertainty. This means that the volume of solution $\Delta V$, which is large and negative at low pressure, approaches the ideal limit ($\Delta V\rightarrow 0$) as pressure increases. The increasing similarity of the volumes of water in the melt and in pure form is consistent with our finding that the water component becomes more structurally interconnected and more like bulk water with increasing pressure. An important consequence of the increasing ideality of the water–silicate solution with increasing pressure is the almost constant density of hydration that we find (Fig. 4a): if $\Delta V=0$ at all pressures, the density contrast between hydrous and anhydrous melts varies rapidly with pressure at low pressure.

We predict that the solubility of water in silicate melt will increase as pressure increases and will be essentially unlimited at all mantle pressure–temperature conditions beyond a few gigapascals. This prediction is based on our finding that the water component behaves increasingly ideally as pressure increases. The volume of solution $\Delta V=\mathrm{d}(\Delta G)/\mathrm{d}P$ is $\leq 0$ at all pressures investigated, so the Gibbs free energy of solution $\Delta G$, which limits solubility, must decrease to a small, pressure-independent value as pressure increases to around 100 GPa, where $\Delta V$ vanishes. As $\Delta G$ is sufficiently small to permit complete solubility at 12 GPa (ref. 18), solubility must remain undiminished as pressure increases to at least the highest pressure we study here (100 GPa). The dependence of $\bar{V}_{H_2O}$ on silicate composition remains to be investigated at high pressure. However, we note that $\bar{V}_{H_2O}$ is apparently very insensitive to silicate composition at low pressure[2], indicating that our arguments regarding solubility may be applicable to natural melt compositions as well.

Essentially unlimited solubility of water in silicate melt over most of the mantle regime would have potentially important implications for our understanding of Earth's origins. In most models of the accreting Earth, a deep magma ocean, possibly encompassing the entire mantle, is an important reservoir of water. Our results show

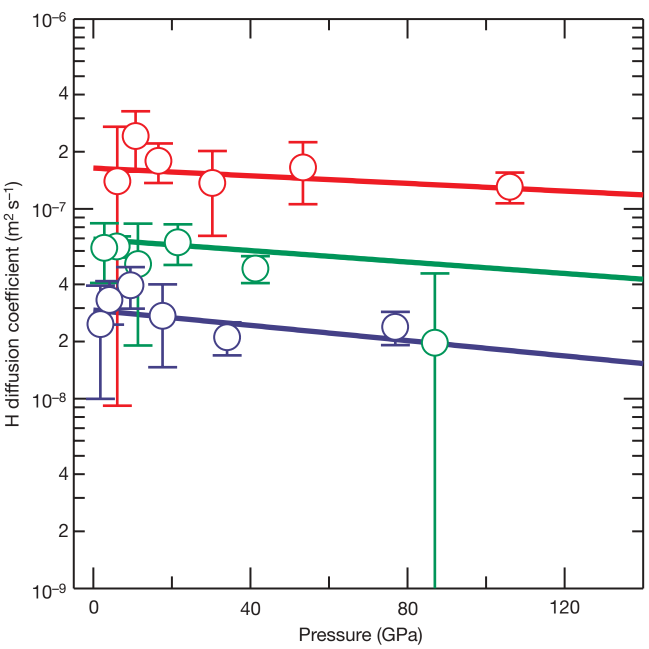



**Figure 3 | Self-diffusion coefficient of hydrogen.** Results at 3,000 K (blue), 4,000 K (green) and 6,000 K (red) from our simulations (symbols) and an Arrhenius fit to the simulation results (lines) (equation (1) with $D_0=9.2\times10^{-7}$ m$^2$ s$^{-1}$, $E^*=86$ kJ mol$^{-1}$ and $V^*=0.11$ cm$^3$ mol$^{-1}$). The value of the activation energy is similar to that found in hydrous basaltic melts at low pressure[29] ($126\pm32$ kJ mol$^{-1}$), and the value of the diffusivity agrees well with that extrapolated to 3,000 K from lower temperature, low-pressure experiments[29] ($2.4\times10^{-8}$ m$^2$ s$^{-1}$), indicating that proton diffusion does not depend strongly on melt composition. Error bars represent one-s.d. uncertainties.

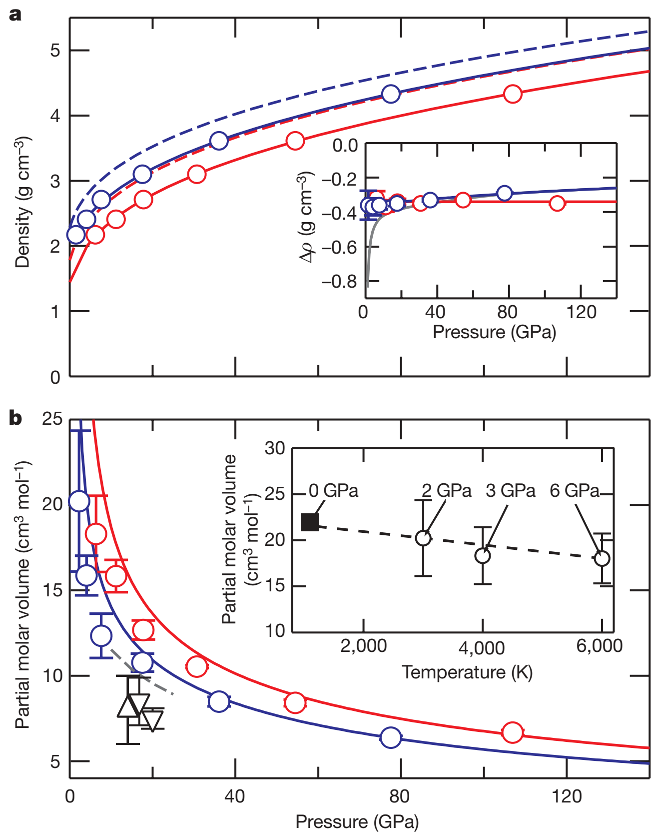



**Figure 4 | Density and partial molar volume of water. a,** Comparison of the density of hydrous (solid lines with symbols) and anhydrous[7] (dashed lines) melts along 3,000 K (blue) and 6,000 K (red) isotherms. The inset shows the density contrast between hydrous and anhydrous melt along 3,000 K (blue) and 6,000 K (red) isotherms. For comparison, the grey line shows the density contrast between an anhydrous melt and an ideal hydrous melt ($\Delta V=0$) at 3,000 K. **b,** Partial molar volume of water from our simulations at 6,000 K (red circles) and 3,000 K (blue circles), compared with that of pure water[30] (solid lines). The partial molar volumes of water derived from high-pressure experiments (upward triangle[16] and downward triangles[17]) are shown for comparison. The dashed line represents the partial molar volume from this study extrapolated to 2,200 K, the experimental temperature in refs 16 and 17. The inset shows the partial molar volume of water (open circles) that we calculate extrapolated along the isochore $V/V_x=1$ to zero pressure (dashed line), in excellent agreement with experiment (solid square[2]). Error bars represent one-s.d. uncertainties.

that this reservoir could easily store all the water delivered to Earth (probably bounded by the 5–7% by mass present in CI carbonaceous chondrites[19]), which probably exceeds the amount of water in the present Earth (0.005–0.02% by mass in the source of mid-ocean-ridge basalts[20]) by a large factor. Moreover, water stored in the magma ocean may have fundamentally altered its thickness, thermal state and dynamics, by means of water-induced lowering of the solidus and the viscosity. Extensive solubility of water in the early magma ocean may have had important implications for the origin of the hydrosphere[21].

## METHODS SUMMARY

The simulations were performed in the canonical ensemble with periodic boundary conditions and a Nosé thermostat[22]. The primary cell contained 84 atoms (12 $MgSiO_3$ molecules and eight water molecules: 10 wt% water). As our initial condition, a pyroxene structure with water molecules placed in the vacant space between tetrahedral chains and in the *b–c* plane was homogeneously strained to a cubic shape and the desired volume. The system was melted at 6,000 K and then cooled isochorically to first 4,000 K and then 3,000 K. We used a time step of 0.5 fs. Total run durations were 3 ps at 6,000 K, 4.5 ps at 4,000 K (except for $V/V_x$ values of 0.5 and 0.6, for which the duration was 6 ps) and 6 ps at 3,000 K (except for a $V/V_x$ value of 0.6, for which the duration was 12.5 ps, and a $V/V_x$ value of 0.5, for which the duration was 15 ps). In all cases the final 80% of the trajectories were used to compute averages.

Self-diffusion coefficients were determined from the slope of the linear portion of the plot of mean squared displacement versus time (see the Supplementary Information). We confirmed that the simulated properties were not significantly affected if we doubled the cell size or used a different initial condition. We used ultrasoft pseudopotentials[23] and VASP[24], computing the electronic structure at the Brillouin zone centre with an energy cutoff of 400 eV. Pulay corrections to the pressure were estimated to be the same as in the anhydrous case[7], varying linearly from 2 GPa at $V = V_x$ to 5 GPa at $V = V_x/2$, where $V_x = 1,033.54$ Å$^3$ per simulation cell (84 atoms), identical to the reference volume of our anhydrous simulations. To account for the error inherent in the approximation to the exchange correlation functional, we added a uniform correction of 2 GPa to the pressure[7]. The simulations were confined to the Born–Oppenheimer surface and include the influence of the finite temperature through the Mermin functional[25,26].

**Received 12 October 2007; accepted 17 March 2008.**


1. Shen, A. H. & Keppler, H. Direct observation of complete miscibility in the albite-$H_2O$ system. *Nature* **385**, 710–712 (1997).
2. Ochs, F. A. & Lange, R. A. The density of hydrous magmatic liquids. *Science* **283**, 1314–1317 (1999).
3. Lange, R. A. The effect of $H_2O$, $CO_2$ and F on the density and viscosity of silicate melts. *Rev. Mineral.* **30**, 331–369 (1994).
4. Stolper, E. The speciation of water in silicate melts. *Geochim. Cosmochim. Acta* **46**, 2609–2620 (1982).
5. Panero, W. R. & Stixrude, L. P. Hydrogen incorporation in stishovite at high pressure and symmetric hydrogen bonding in delta-AlOOH. *Earth Planet. Sci. Lett.* **221**, 421–431 (2004).
6. Pohlmann, M., Benoit, M. & Kob, W. First-principles molecular-dynamics simulations of a hydrous silica melt: Structural properties and hydrogen diffusion mechanism. *Phys. Rev. B* **70**, 184209 (2004).
7. Stixrude, L. & Karki, B. Structure and freezing of $MgSiO_3$ liquid in Earth's lower mantle. *Science* **310**, 297–299 (2005).
8. Closmann, C. & Williams, Q. In-situ spectroscopic investigation of high-pressure hydrated (Mg,Fe)$SiO_3$ glasses - OH vibrations as a probe of glass structure. *Am. Mineral.* **80**, 201–212 (1995).
9. Inoue, T. Effect of water on melting phase relations and melt composition in the system $Mg_2SiO_4$-$MgSiO_3$-$H_2O$ up to 15 GPa. *Phys. Earth Planet. Inter.* **85**, 237–263 (1994).
10. Goldman, N., Fried, L. E., Kuo, I. F. W. & Mundy, C. J. Bonding in the superionic phase of water. *Phys. Rev. Lett.* **94**, 217801 (2005).
11. Greaves, G. N. & Ngai, K. L. Reconciling ionic-transport properties with atomic-structure in oxide glasses. *Phys. Rev. B* **52**, 6358–6380 (1995).
12. Song, T. R. A., Helmberger, D. V. & Grand, S. P. Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States. *Nature* **427**, 530–533 (2004).
13. Toffelmier, D. A. & Tyburczy, J. A. Electromagnetic detection of a 410-km-deep melt layer in the southwestern United States. *Nature* **447**, 991–994 (2007).
14. Dziewonski, A. M. & Anderson, D. L. Preliminary reference earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
15. Ito, E. & Takahashi, E. Melting of peridotite at uppermost lower-mantle conditions. *Nature* **328**, 514–517 (1987).
16. Matsukage, K. N., Jing, Z. C. & Karato, S. Density of hydrous silicate melt at the conditions of Earth's deep upper mantle. *Nature* **438**, 488–491 (2005).
17. Sakamaki, T., Suzuki, A. & Ohtani, E. Stability of hydrous melt at the base of the Earth's upper mantle. *Nature* **439**, 192–194 (2006).
18. Stalder, R., Ulmer, P., Thompson, A. B. & Gunther, D. High pressure fluids in the system MgO-$SiO_2$-$H_2O$ under upper mantle conditions. *Contrib. Mineral. Petrol.* **140**, 607–618 (2001).
19. Kerridge, J. F. Carbon, hydrogen and nitrogen in carbonaceous chondrites - abundances and isotopic compositions in bulk samples. *Geochim. Cosmochim. Acta* **49**, 1707–1714 (1985).
20. Hirschmann, M. M. Water, melting, and the deep Earth $H_2O$ cycle. *Annu. Rev. Earth Planet. Sci.* **34**, 629–653 (2006).
21. Matsui, T. & Abe, Y. Evolution of an impact-induced atmosphere and magma ocean on the accreting Earth. *Nature* **319**, 303–305 (1986).
22. Nosé, S. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
23. Kresse, G. & Hafner, J. Norm-conserving and ultrasoft pseudopotentials for first-row and transition-elements. *J. Phys. Condens. Matter* **6**, 8245–8257 (1994).
24. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
25. Mermin, N. D. Thermal properties of inhomogeneous electron gas. *Phys. Rev.* **137**, A1441–A1443 (1965).
26. Wentzcovitch, R. M., Martins, J. L. & Allen, P. B. Energy versus free-energy conservation in 1st-principles molecular-dynamics. *Phys. Rev. B* **45**, 11372–11374 (1992).
27. Zhang, Y. X. $H_2O$ in rhyolitic glasses and melts: Measurement, speciation, solubility, and diffusion. *Rev. Geophys.* **37**, 493–516 (1999).
28. Bhattarai, D., Karki, B. B. & Stixrude, L. Space-time multiresolution atomistic visualization of MgO and $MgSiO_3$ liquid data. *Vis. Geosci.* **11**, 1–11 (2006).
29. Zhang, Y. X. & Stolper, E. M. Water diffusion in a basaltic melt. *Nature* **351**, 306–309 (1991).
30. Pitzer, K. S. & Sterner, S. M. Equations of state valid continuously from zero to extreme pressures for $H_2O$ and $CO_2$. *J. Chem. Phys.* **101**, 3111–3116 (1994).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** Authors thank the Center for Computation & Technology at Louisiana State University for computing resources. This work was supported by the US National Science Foundation.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.S. (stixrude@umich.edu).

LETTERS

# Mountain pine beetle and forest carbon feedback to climate change

W. A. Kurz[1], C. C. Dymond[1], G. Stinson[1], G. J. Rampley[1], E. T. Neilson[1], A. L. Carroll[1], T. Ebata[2] & L. Safranyik[1]

**The mountain pine beetle (*Dendroctonus ponderosae* Hopkins, Coleoptera: Curculionidae, Scolytinae) is a native insect of the pine forests of western North America, and its populations periodically erupt into large-scale outbreaks[1–3]. During outbreaks, the resulting widespread tree mortality reduces forest carbon uptake and increases future emissions from the decay of killed trees. The impacts of insects on forest carbon dynamics, however, are generally ignored in large-scale modelling analyses. The current outbreak in British Columbia, Canada, is an order of magnitude larger in area and severity than all previous recorded outbreaks[4]. Here we estimate that the cumulative impact of the beetle outbreak in the affected region during 2000–2020 will be 270 megatonnes (Mt) carbon (or 36 g carbon $m^{-2} yr^{-1}$ on average over 374,000 $km^2$ of forest). This impact converted the forest from a small net carbon sink to a large net carbon source both during and immediately after the outbreak. In the worst year, the impacts resulting from the beetle outbreak in British Columbia were equivalent to ~75% of the average annual direct forest fire emissions from all of Canada during 1959–1999. The resulting reduction in net primary production was of similar magnitude to increases observed during the 1980s and 1990s as a result of global change[5]. Climate change has contributed to the unprecedented extent and severity of this outbreak[6]. Insect outbreaks such as this represent an important mechanism by which climate change may undermine the ability of northern forests to take up and store atmospheric carbon, and such impacts should be accounted for in large-scale modelling analyses.**

Forest insect epidemics can have severe impacts on ecosystem dynamics by causing mortality and reducing the growth of millions of trees over extensive areas[7]. Native insects and alien invasive species affect both managed and natural forests. Beyond the ecological impacts are the associated economic (for example, disrupted timber supply to mills) and social (for example, unemployment, crime rates) effects[8]. The impact of insects on carbon (C) dynamics and global climate are not well documented[9].

The current outbreak of mountain pine beetle in western Canada is an order of magnitude greater in area than previous outbreaks owing to the increased area of susceptible host (mature pine stands) and favourable climate[4] (see also Supplementary Fig. 3). An expansion in climatically suitable habitat for the mountain pine beetle, including reduced minimum winter temperature, increased summer temperatures and reduced summer precipitation, during recent decades has facilitated expansion of the outbreak northward and into higher elevation forests[4,10]. This range expansion, combined with an increase in the extent of the host, has resulted in an outbreak of unprecedented scale and severity. By the end of 2006, the cumulative outbreak area was 130,000 $km^2$ (many stands being attacked in multiple years), with tree mortality ranging from single trees to most of a stand in a single year[11]. Timber losses are estimated to be more than 435 million $m^3$, with additional losses outside the commercial forest[12]. The forest sector has responded by increasing harvest rates and reallocating some harvest, increasing the pine portion of the provincial total volume harvested from 31% to 45% over four years (2001–2004).

We estimated the combined impact of the beetle, forest fires and harvesting on forest productivity and carbon balance from 2000 until 2020 for the south-central region of British Columbia (Fig. 1). This area includes 374,000 $km^2$ of productive forest, largely dominated by pine (*Pinus*) and spruce (*Picea*) species. We used a Monte Carlo design for simulating future net biome production (NBP) using a forest ecosystem model (the Carbon Budget Model of the Canadian Forest Sector, CBM-CFS3). This model accounts for annual tree growth, litterfall, turnover and decay, and it explicitly simulates harvest, beetle-caused mortality, and fire-caused mortality and fuel consumption. We developed regional probability distribution functions of the annual area burned and projected future beetle dynamics on the basis of the characteristics of the remaining host (that is, pine stands of suitable age) and the judgement of regional entomologists. We conducted 100 Monte Carlo simulations with different random draws from the probability distributions for the annual area of beetle outbreak and the annual area burned.

For the period 2000–2020, the average annual NBP was $-15.8 \pm 7.9$ Mt C $yr^{-1}$ (or $-42.4 \pm 21$ g C $m^{-2} yr^{-1}$; Fig. 2). Carbon losses result from emissions from decomposition and fires and from the transfer of timber to the forest product sector. In a separate analysis[13], we estimated that the study area was a net sink from 1990 to 2002. The first two years of this study also reported a net sink (0.59 Mt C $yr^{-1}$), but with increasing beetle impact (Fig. 3), the forest converted to a source of 17.6 Mt C $yr^{-1}$from 2003 to 2020. With decreasing beetle impact (Fig. 3), NBP began to recover, but by 2020, the estimated NBP had not yet returned to pre-outbreak levels. Although we can expect that forests will eventually recover from the beetle outbreak, we are reluctant to extend projections beyond 2020 or to speculate on the rate of recovery beyond 2020 given uncertainties about non-host responses, rates of regeneration, and future fires in a region in which major climate change impacts are forecast[14].

One component of the uncertainty in future NBP is that we do not know the future area that will be infested by the beetle. We projected the area infested during 2007–2020 using random draws from regionally calibrated probability distributions of outbreak area and duration that were based on: the 2000–2006 area; mortality and host statistics; historical, spatial and temporal dynamics; remaining host population; and judgment from entomologists. The outbreak was projected to peak between 2006 and 2008, with the maximum area infested ranging from 74,000 $km^2$ to 94,000 $km^2$ (Fig. 3).

[1]Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, V8Z 1M5, Canada. [2]British Columbia Ministry of Forests and Range, Victoria, British Columbia, V8W 9C2, Canada.

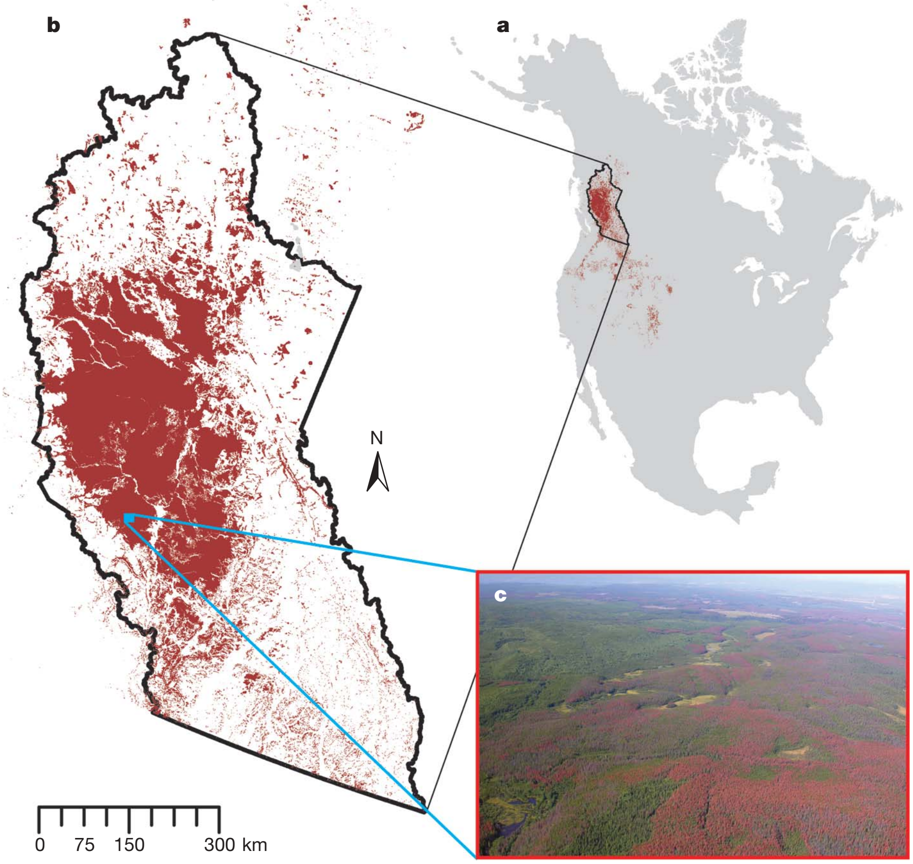



**Figure 1 | Geographic extent of mountain pine beetle outbreak in North America.** **a**, Extent (dark red) of mountain pine beetle. **b**, The study area includes 98% of the current outbreak area. **c**, A photograph taken in 2006 showing an example of recent mortality: pine trees turn red in the first year after beetle kill, and grey in subsequent years. Photo credit: Joan Westfall, Entopath Management Ltd.

We partitioned the impact of the beetle outbreak on NBP using a single projection of beetle (Fig. 3) and fire area and by simulating three forest carbon dynamics scenarios: without beetle (control), with beetle but no additional harvest, and with beetle and additional harvest. The additional harvest is considered to be part of the beetle impact because it is a management response to the vast quantities of beetle-killed trees. The impact of beetles on NBP was projected to reach $-20\,\mathrm{Mt\,C\,yr^{-1}}$ ($-53\,\mathrm{g\,C\,m^{-2}\,yr^{-1}}$) in 2009 (Fig. 4). Cumulative beetle impact over the 21-yr simulation period was projected to be $-270$ Mt C, and the additional salvage harvest and removal of both beetle-killed and living trees contributed a further $-50$ Mt C. Cumulative transfer of beetle-killed timber to the forest product sector was projected to be 31 Mt C (13% of total harvest). Harvesting results in a loss of carbon from the ecosystem, but only some of that carbon is emitted to the atmosphere; the remainder is stored in wood products and landfills.

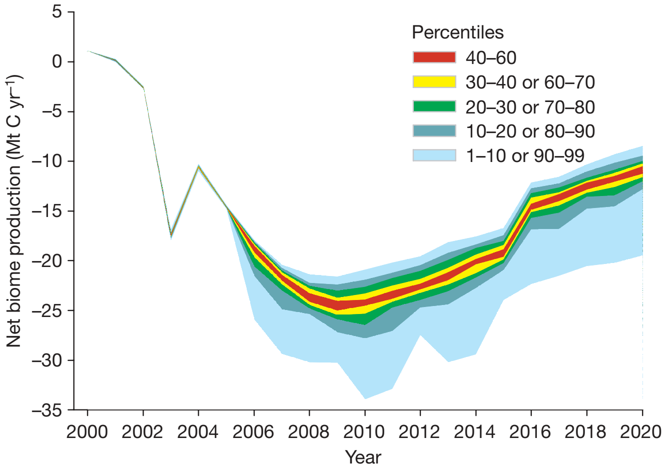



**Figure 2 | Annual NBP by percentile from the Monte Carlo simulations.** The model results are based on statistics from 2000 to 2006 and projections from 2007 to 2020. Negative values represent fluxes from the forest to the atmosphere (a net source of carbon). Asymmetry of the range of estimates of NBP in any single year is a function of the area burned and the associated direct carbon emissions.

NBP was reduced in the affected landscape by the beetle and the associated additional harvest response because of reduced net primary production (NPP) and increased heterotrophic respiration. Annual NPP dropped from 440 to $400\,\mathrm{g\,C\,m^{-2}}$ in 2000–2009. The lowest NPP values ($391\,\mathrm{g\,C\,m^{-2}}$) were estimated for 2015 to 2018, at which point NPP started to recover. NPP was reduced in the affected landscape during and after the beetle outbreak because of the reduction in photosynthetic capacity caused by widespread tree mortality. Conversely, heterotrophic respiration increased from 408 to $424\,\mathrm{g\,C\,m^{-2}}$ in the 2003–2007 period because of the large transfers of biomass to dead organic matter pools (471 Mt C transferred by the beetle during the outbreak) and subsequent decomposition.

Impacts of the current beetle outbreak are important because they are of comparable magnitude to other global change factors. Fire emissions vary widely between years, but within the study area the worst year for fires in the simulation period produced direct emissions of 13 Mt C, an amount exceeded by annual beetle-caused impacts from 2005 to 2014. The maximum annual beetle impact ($20\,\mathrm{Mt\,C\,yr^{-1}}$) for the relatively small study area is of similar magnitude to the direct forest fire emissions from all of Canada during 1959–1999 ($27\,\mathrm{Mt\,C\,yr^{-1}}$)[15]. Climate warming, elevated atmospheric $CO_2$ concentrations, and atmospheric nitrogen deposition may

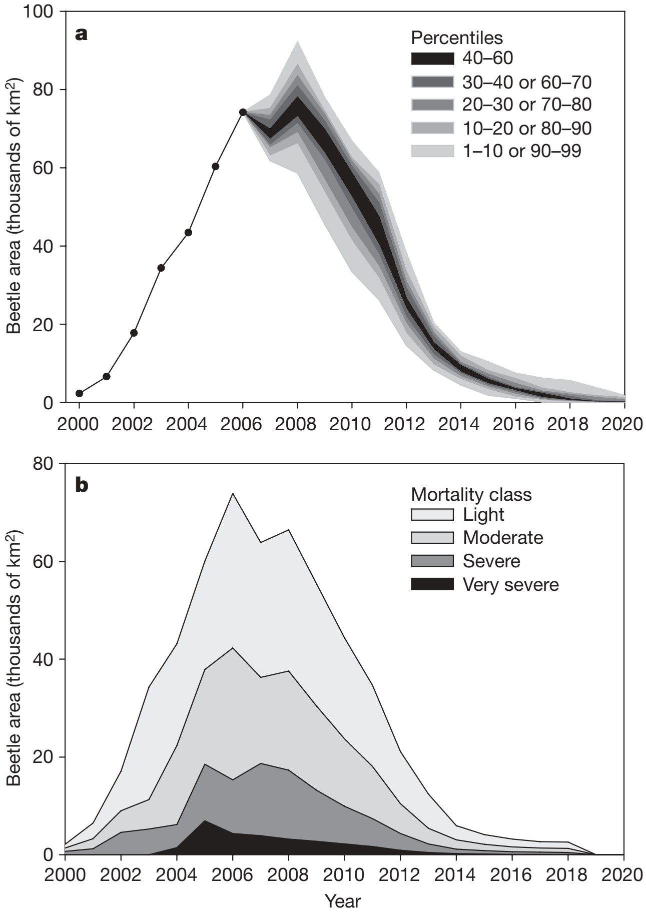



**Figure 3 | Area infested with the mountain pine beetle during the simulation period.** Statistics are used from 2000 to 2006 and projections are used from 2007 to 2020. **a**, Percentiles describing the parameter space resulting from 100 Monte Carlo projections of the beetle-infested area. Our projections of decreasing area after 2009 are largely based on the decline in available live host (pine) area. **b**, Area statistics for the single Monte Carlo simulation used for scenario analysis broken down by host mortality class.

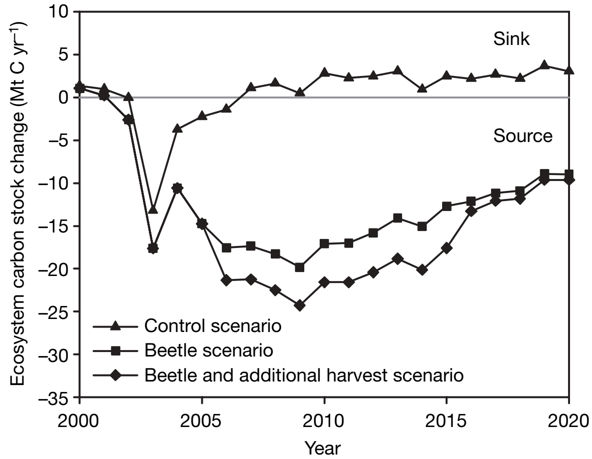



**Figure 4 | Total ecosystem carbon stock change for three scenarios.** The control simulation was run with no beetle outbreak, and with base harvest and fires. The beetle simulation added insect impacts to the control scenario. The additional harvest simulation added the management response of increased harvest levels from 2006 to 2016 to the beetle simulation. Negative ecosystem carbon stock change values represent fluxes from the forest to the atmosphere (net source of carbon). The source in 2003 was, in part, the result of the large area burned (2,440 km$^2$ in the study area) that was included in all three scenarios.

enhance forest carbon sinks[16,17]. For example, it was estimated in ref. 5 that NPP increased 6% during the 1980s and 1990s as a result of global change. The impact of the beetle outbreak would negate these gains within the affected region because NPP was reduced by 10% and heterotrophic respiration increased by 6%. Here we examined just one insect outbreak and estimated that its net greenhouse gas impacts over 21 yr (990 Mt carbon dioxide equivalents, $CO_2$e) are comparable to about 5 yr of emissions from Canada's transportation sector (200 Mt $CO_2$e in 2005)[13].

Climate change will have impacts on insect distribution and abundance either directly via effects on their life cycles or indirectly via host–plant defences, the abundance of natural enemies or interactions with competitors[18]. For some forest insects, the outcome may be changes in: outbreak frequencies or duration; rates of herbivory and damage; ranges, and associations with host species [11,19]. Evidence for climate-change-related increases in the extent or severity of forest insect disturbance has begun to accumulate. In north-western North America, warmer temperatures have halved the time required to reproduce for the spruce beetle (*Dendroctonus rufipennis* Kirby, Coleoptera: Curculionidae, Scolytinae) and have contributed to unprecedented damage to spruce forests[20]. In the western United States, warmer annual temperatures have caused an altitudinal shift in the range of habitats, allowing the mountain pine beetle to invade high-elevation pine forests[21]. Insect impacts have, however, generally been ignored in large-scale carbon budget modelling studies[22,23] because data on insect impacts covering large areas are expensive to collect and are therefore not widely available[24]. Moreover, complex modelling approaches are required to adequately represent the impacts of insects on stand- and landscape-level forest ecosystem dynamics. This failure to account for insect impacts may have resulted in overestimation in previous studies of the potential for forests to offset anthropogenic $CO_2$.

Other insects, such as eastern spruce budworm (*Choristoneura fumiferana* Clemens, Lepidoptera: Tortricidae) and forest tent caterpillar (*Malacosoma disstria* Hubner, Lepidoptera: Lasiocampidae) also exhibit outbreaks during which they affect forest carbon dynamics by reducing growth and increasing tree mortality over large areas[25]. Other forms of disturbance, such as fires, also markedly influence terrestrial carbon budgets[26]. Significant climate warming has already allowed the mountain pine beetle to expand its range into formerly climatically unsuitable habitats[4,21]. Future projected warming could facilitate further range expansion. Insect outbreaks, together with fires, could put North American forest carbon sinks at risk[27]. Disturbances are one of the principal drivers of the carbon budget in northern forests[26], and this study shows but one example of how climate change can affect disturbance regimes for which impacts on the global carbon cycle then provide strong positive feedback to the global climate system.

## METHODS SUMMARY

The CBM-CFS3 uses forest inventory as well as growth and yield data to simulate stand- and landscape-level forest carbon dynamics[28–30]. It tracks carbon stocks, transfers between pools and emissions. The model accounts for tree growth, litterfall, turnover and decay as annual components of the forest carbon cycle. Fires were simulated as stand-replacing events where a portion of foliage and above-ground dead organic matter pools were consumed; all remaining biomass was killed, and standing snags were transferred to litter and coarse woody debris. Harvests were also simulated as stand-replacing clear-cut events where stemwood carbon was transferred out of the ecosystem to the forest products sector, and the rest of the carbon in biomass pools was transferred to pools of dead organic matter. Beetle infestation areas were simulated as partial-mortality events killing a portion of softwood biomass pools. The remaining stand continues to grow and recover from beetle impacts. The stand mortality classes were 5%, 10%, 30% or 50% per year (see Supplementary Information for more details).

The British Columbia Ministry of Forests and Range supplied forest inventory and growth curve data. Almost 700,000 forest stands were assigned to 86 spatial analysis units that represented the administrative and ecological variability in the study area. Spatially and temporally explicit beetle outbreak area and

area-burned data sets were used for 2000–2006 and projected for 2007–2020 using random draws from historically calibrated probability distributions. Harvest rates for each administrative unit were available for 2000–2005, and forecasts for 2006–2020 were based on local management plans (see Supplementary Information for more details).

**Received 9 December 2007; accepted 29 January 2008.**

1. Amman, G. D. & Cole, W. E. in *Mountain Pine Beetle Dynamics in Lodgepole Pine Forests Part II: Population Dynamics.* General Technical Report No. GTR-INT-145 (United States Department of Agriculture Forest Service, Intermountain Forest and Range Experiment Station, Ogden, 1983).
2. Safranyik, L., Shrimpton, D. & Whitney, H. in *Management of Lodgepole Pine to Reduce Losses from the Mountain Pine Beetle. Technical Report 1* (Environment Canada, Canadian Forestry Service, Victoria, 1974).
3. Taylor, S. W. & Carroll, A. L. in *Mountain Pine Beetle Symposium: Challenges and Solutions BC-X-399* (eds Shore, T. L., Brooks, J. E. & Stone, J. E.) 41–51 (Natural Resources Canada, Canadian Forest Service, Victoria, 2004).
4. Taylor, S. W., Carroll, A. L., Alfaro, R. I. & Safranyik, L. in *The Mountain Pine Beetle: A Synthesis of Biology, Management and Impacts in Lodgepole Pine* (eds Safranyik, L. & Wilson, B.) 67–94 (Natural Resources Canada, Canadian Forest Service, Victoria, 2006).
5. Nemani, R. R. *et al.* Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300,** 1560–1563 (2003).
6. Carroll, A. L., Taylor, S. W., Regniere, J. & Safranyik, L. in *Mountain Pine Beetle Symposium: Challenges and Solutions*. Report number BC-X-399 (eds Shore, T. L., Brooks, J. E. & Stone, J. E.) 223–232 (Natural Resources Canada, Canadian Forest Service, Victoria, 2004).
7. Mattson, W. J. & Addy, N. D. Phytophagous insects as regulators of forest primary productivity. *Science* **190,** 515–522 (1975).
8. Parkins, J. R. & McKendrick, N. A. Assessing community vulnerability: a study of the mountain pine beetle outbreak in British Columbia, Canada. *Glob. Environ. Chang.* **17,** 460–471 (2007).
9. Volney, W. J. A. & Fleming, R. A. Climate change and impacts of boreal forest insects. *Agr. Ecosys. Env.* **82,** 283–294 (2000).
10. Williams, D. W. & Liebhold, A. M. Climate change and the outbreak ranges of two North American bark beetles. *Agr. For. Ent.* **4,** 87–99 (2002).
11. Westfall, J. *2006 Summary of Forest Health Conditions in British Columbia* (British Columbia Ministry of Forests and Range, Victoria, 2007).
12. Walton, A. *et al. Provincial-Level Projection of the Current Mountain Pine Beetle Outbreak* ⟨http://www.for.gov.bc.ca/hre/bcmpb/Year4.htm⟩ (British Columbia, Ministry of Forest and Range, Victoria, 2007).
13. Environment Canada. *National Inventory Report, 1990–2005: Greenhouse Gas Sources and Sinks in Canada.* ⟨http://www.ec.gc.ca/pdb/ghg/inventory_report/2005_report/tdm-toc_eng.cfm⟩. (Environment Canada, Gatineau, 2007).
14. Hamann, A. & Wang, T. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology* **87,** 2773–2786 (2006).
15. Amiro, B. D. *et al.* Direct carbon emissions from Canadian forest fires, 1959–1999. *Can. J. For. Res.* **31,** 512–525 (2001).
16. Canadell, J. G. *et al.* in *Terrestrial Ecosystems in a Changing World* (eds Canadell, J. G., Pataki, D. E. & Pitkelka, L. F.) 59–78 (Springer, Berlin, 2007).
17. Denman, K. L. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) 499–587 (Cambridge Univ. Press, Cambridge, 2007).
18. Ayres, M. P. & Lombardero, M. J. Assessing the consequences of global change for forest disturbance from herbivores and pathogens. *Sci. Total Environ.* **262,** 263–286 (2000).
19. Kurz, W. A., Apps, M. J., Stocks, B. J. & Volney, W. J. A. in *Biotic Feedbacks in the Global Climatic System: Will the Warming Feed the Warming?* (eds Woodwell, G. M. & Mackenzie, F. T.) 119–133 (Oxford Univ. Press, Oxford, 1995).
20. Berg, E. E., Henry, J. D., Fastie, C. L., De Volder, A. D. & Matsuoka, S. M. Spruce beetle outbreaks on the Kenai Peninsula, Alaska, and Kluane National Park and Reserve, Yukon Territory: relationship to summer temperatures and regional differences in disturbance regimes. *For. Ecol. Manage.* **227,** 219–232 (2006).
21. Logan, J. A. & Powell, J. A. Ghost forests, global warming, and the mountain pine beetle (Coleoptera: Scolytidae). *Am. Ent.* **47,** 160–173 (2001).
22. McGuire, A. D. *et al.* Carbon balance of the terrestrial biosphere in the twentieth century: analyses of $CO_2$, climate and land use effects with four process-based ecosystem models. *G. Biogeoch. Cycles* **15,** 183–206 (2001).
23. Myneni, R. B. *et al.* A large carbon sink in the woody biomass of northern forests. *Proc. Nat. Acad. Sci.* **98,** 14784–14789 (2001).
24. Wulder, M. A., Dymond, C. C., White, J. C., Leckie, D. G. & Carroll, A. L. Surveying mountain pine beetle damage of forests: a review of remote sensing opportunities. *For. Ecol. Manage.* **221,** 27–41 (2001).
25. Berryman, A. A. *Dynamics of Forest Insect Populations: Patterns, Causes, Implications* (Plenum, New York, 1988).
26. Bond-Lamberty, B., Peckham, S. D., Ahl, D. E. & Gower, S. T. Fire as the dominant driver of central Canadian boreal forest carbon balance. *Nature* **450,** 89–92 (2007).
27. Kurz, W. A., Stinson, G. & Rampley, G. Could increased boreal forest ecosystem productivity offset carbon losses from increased disturbances? *Phil. Trans. R. Soc. B.*. doi:10.1098/rstb.2007.2198 (2007).
28. Kurz, W. A., Apps, M. J., Webb, T. M. & McNamee, P. J. *The Carbon Budget of the Canadian Forest Sector: Phase I.* Information Report NOR-X-326 (Forestry Canada, Northwest Region, Northern Forestry Centre, Edmonton, 1992).
29. Kurz, W. A. & Apps, M. J. A 70-year retrospective analysis of carbon fluxes in the Canadian forest sector. *Ecol. Appl.* **9,** 526–547 (1999).
30. Kull, S. *et al. Operational-Scale Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3) Version 1.0: User's Guide* (Natural Resources Canada, Canadian Forest Service, Edmonton, 2006).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** Funding for this study was provided by Canada's Climate Change Action Plan. We thank British Columbia Ministry of Forests and Range (MoFR) collaborators for providing data and detailed reviews: D. Draper, G. Lawrence, M. Boyce and D. Spittlehouse. S. Beukema provided assistance in formatting MoFR data for use in CBM-CFS3. Fire and beetle projection methodologies were developed with the assistance of J. Metsaranta, S. Beukema, B. Stocks, B. Amiro, M. Flannigan, R. Landry, B. de Groot, K. Anderson, S. Taylor and T. Shore. Climate data were provided by D. McKenney. CBM-CFS3 model development was assisted by T. White, G. Zhang, M. Magnan, E. Banfield, C. Shaw and B. Simpson. Finally, we thank V. Nealis, D. Draper, D. Spittlehouse and A. Nussbaum for comments on an earlier draft.

**Author Contributions** All authors contributed to study design, model parameterizations, and development of future projections; C.C.D., G.J.R., G.S. and E.T.N. conducted model simulations; W.A.K., C.C.D., G.S. and E.T.N. analysed results; and all authors contributed to the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.A.K. (wkurz@nrcan.gc.ca).

nature

LETTERS

# The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus)

Ray Ming[1,2]*, Shaobin Hou[3]*, Yun Feng[4,5]*, Qingyi Yu[1]*, Alexandre Dionne-Laporte[3], Jimmy H. Saw[3], Pavel Senin[3], Wei Wang[4,6], Benjamin V. Ly[3], Kanako L. T. Lewis[3], Steven L. Salzberg[7], Lu Feng[4,5,6], Meghan R. Jones[1], Rachel L. Skelton[1], Jan E. Murray[1,2], Cuixia Chen[2], Wubin Qian[4], Junguo Shen[5], Peng Du[5], Moriah Eustice[1,8], Eric Tong[1], Haibao Tang[9], Eric Lyons[10], Robert E. Paull[11], Todd P. Michael[12], Kerr Wall[13], Danny W. Rice[14], Henrik Albert[15], Ming-Li Wang[1], Yun J. Zhu[1], Michael Schatz[7], Niranjan Nagarajan[7], Ricelle A. Acob[1,8], Peizhu Guan[1,8], Andrea Blas[1,8], Ching Man Wai[1,11], Christine M. Ackerman[1], Yan Ren[4], Chao Liu[4], Jianmei Wang[4], Jianping Wang[2], Jong-Kuk Na[2], Eugene V. Shakirov[16], Brian Haas[17], Jyothi Thimmapuram[18], David Nelson[19], Xiyin Wang[9], John E. Bowers[9], Andrea R. Gschwend[2], Arthur L. Delcher[7], Ratnesh Singh[1,8], Jon Y. Suzuki[15], Savarni Tripathi[15], Kabi Neupane[20], Hairong Wei[21], Beth Irikura[11], Maya Paidi[1,8], Ning Jiang[22], Wenli Zhang[23], Gernot Presting[8], Aaron Windsor[24], Rafael Navajas-Pérez[9], Manuel J. Torres[9], F. Alex Feltus[9], Brad Porter[8], Yingjun Li[2], A. Max Burroughs[7], Ming-Cheng Luo[25], Lei Liu[18], David A. Christopher[8], Stephen M. Mount[7,26], Paul H. Moore[15], Tak Sugimura[27], Jiming Jiang[23], Mary A. Schuler[28], Vikki Friedman[29], Thomas Mitchell-Olds[24], Dorothy E. Shippen[16], Claude W. dePamphilis[13], Jeffrey D. Palmer[14], Michael Freeling[10], Andrew H. Paterson[9], Dennis Gonsalves[15], Lei Wang[4,5,6] & Maqsudul Alam[3,30]

**Papaya, a fruit crop cultivated in tropical and subtropical regions, is known for its nutritional benefits and medicinal applications. Here we report a 3× draft genome sequence of 'SunUp' papaya, the first commercial virus-resistant transgenic fruit tree[1] to be sequenced. The papaya genome is three times the size of the *Arabidopsis* genome, but contains fewer genes, including significantly fewer disease-resistance gene analogues. Comparison of the five sequenced genomes suggests a minimal angiosperm gene set of 13,311. A lack of recent genome duplication, atypical of other angiosperm genomes sequenced so far[2–5], may account for the smaller papaya gene number in most functional groups. Nonetheless, striking amplifications in gene number within particular functional groups suggest roles in the evolution of tree-like habit, deposition and remobilization of starch reserves, attraction of seed dispersal agents, and adaptation to tropical daylengths. Transgenesis at three locations is closely associated with chloroplast insertions into the nuclear genome, and with topoisomerase I recognition sites. Papaya offers numerous advantages as a system for fruit-tree functional genomics, and this draft genome sequence provides the foundation for revealing the basis of *Carica*'s distinguishing morpho-physiological, medicinal and nutritional properties.**

Papaya is an exceptionally promising system for the exploration of tropical-tree genomes and fruit-tree genomics. It has a relatively small genome of 372 megabases (Mb)[6], diploid inheritance with nine pairs of chromosomes, a well-established transformation system[7], a short generation time (9–15 months), continuous flowering throughout the year and a primitive sex-chromosome system[8]. It is a member of the Brassicales, sharing a common ancestor with *Arabidopsis* about 72 million years ago[9]. Papaya is ranked first on nutritional scores among 38 common fruits, based on the percentage of the United States Recommended Daily Allowance for vitamin A, vitamin C, potassium, folate, niacin, thiamine, riboflavin, iron and calcium, plus fibre. Consumption of its fruit is recommended for preventing vitamin A deficiency, a cause of childhood blindness in tropical and subtropical developing countries. The fruit, stems, leaves and roots of papaya are used in a wide range of medical applications, including production of papain, a valuable proteolytic enzyme.

[1]Hawaii Agriculture Research Center, Aiea, Hawaii 96701, USA. [2]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3]Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii 96822, USA. [4]TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin Economic-Technological Development Area, Tianjin 300457, China. [5]Tianjin Research Center for Functional Genomics and Biochip, Tianjin Economic-Technological Development Area, Tianjin 300457, China. [6]Key Laboratory of Molecular Microbiology and Technology of the Ministry of Education, College of Life Sciences, Nankai University, Tianjin 300071, China. [7]Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland 20742, USA. [8]Department of Molecular Bioscience and Bioengineering, University of Hawaii, Honolulu, Hawaii 96822, USA. [9]Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30602, USA. [10]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. [11]Department of Tropical Plant and Soil Sciences, University of Hawaii, Honolulu, Hawaii 96822, USA. [12]Waksman Institute of Microbiology and Department of Plant Biology and Pathology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA. [13]Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. [14]Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. [15]USDA-ARS, Pacific Basin Agricultural Research Center, Hilo, Hawaii 96720, USA. [16]Department of Biochemistry and Biophysics, 2128 TAMU, Texas A&M University, College Station, Texas 77843, USA. [17]The Institute for Genomic Research, Rockville, Maryland 20850, USA. [18]W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [19]Department of Molecular Sciences, University of Tennessee, Memphis, Tennessee 38163, USA. [20]Leeward Community College, University of Hawaii, Pearl City, Hawaii 96782, USA. [21]Wicell Research Institute, Madison, Wisconsin 53707, USA. [22]Department of Horticulture, Michigan State University, East Lansing, Michigan 48824, USA. [23]Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706, USA. [24]Department of Biology, Duke University, Durham, North Carolina 27708, USA. [25]Department of Plant Sciences, University of California, Davis, California 95616, USA. [26]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland 20742, USA. [27]Maui High Performance Computing Center, Kihei, Hawaii 96753, USA. [28]Departments of Cell and Developmental Biology, Biochemistry and Plant Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [29]Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA. [30]Department of Microbiology, University of Hawaii, Honolulu, Hawaii 96822, USA.
*These authors contributed equally to this work.

A total of 2.8 million whole-genome shotgun (WGS) sequencing reads were generated from a female plant of transgenic cultivar SunUp, which was developed through transformation of Sunset that had undergone more than 25 generations of inbreeding[10]. The estimated residual heterozygosity of SunUp is 0.06% (Supplementary Note 1). After excluding low-quality and organellar reads, 1.6 million high-quality reads were assembled into contigs containing 271 Mb and scaffolds spanning 370 Mb including embedded gaps (Supplementary Tables 1 and 2). Of 16,362 unigenes derived from expressed sequence tags (ESTs), 15,064 (92.1%) matched this assembly. Paired-end reads from 34,065 bacterial artificial chromosome (BAC) clones provided alignment to an fingerprinted contig (FPC)-based physical map (Supplementary Note 2). Among 706 BAC end and WGS sequence-derived simple sequence repeats on the genetic map, 652 (92.4%) could be used to anchor 167 Mb of contigs or 235 Mb of scaffolds, to the 12 papaya linkage groups in the current genetic map (Supplementary Fig. 1).

Papaya chromosomes at the pachytene stage of meiosis are generally stained lightly by 4′,6-diamidino-2-phenylindole (DAPI), revealing that the papaya genome is largely euchromatic. However, highly condensed heterochromatin knobs were observed on most chromosomes (Supplementary Fig. 2), concentrated in the centromeric and pericentromeric regions. The lengths of the pachytene bivalents that are heavily stained only account for approximately 17% of the genome. However, these cytologically distinct and highly condensed heterochromatic regions could represent 30–35% of the genomic DNA[11]. A large portion of the heterochromatic DNA was probably not covered by the WGS sequence. The 271 Mb of contig sequence should represent about 75% of the papaya genome and more than 90% of the euchromatic regions, which is similar to the 92.1% of the EST and 92.4% of genetic markers covered by the assembled genome and the theoretical 95% coverage by 3× WGS sequence[12].

Gene annotation was carried out using the TIGR Eukaryotic Annotation Pipeline. The assembled genome was masked based on similarity to known repeat elements in RepBase and the TIGR Plant Repeat Database, plus a *de novo* papaya repeat database (see Methods). *Ab initio* gene predictions were combined with spliced alignments of proteins and transcripts to produce a reference gene set of 28,629 gene models (Supplementary Table 3). A total of 21,784 (76.1%) of the predicted papaya genes with average length of 1,057 base pairs (bp) have similarity to proteins in the non-redundant database from the National Center for Biotechnology Information, with 9,760 (44.8%) of these supported by papaya unigenes. Among 6,845 genes with average length 309 bp that had no hits to the non-redundant proteins, only 515 (7.5%) were supported by papaya unigenes, implying that the number of predicted papaya-specific genes was inflated. If the 515 genes with unigene support represent 44.8% of the total, then 1,150 predicted papaya-specific genes may be real, and the number of predicted genes in the assembled papaya genome would be 22,934. Considering the assembled genome covers 92.1% of the unigenes and 92.4% of the mapped genetic markers, the number of predicted genes in the papaya genome could be 7.9% higher, or 24,746, about 11–20% less than *Arabidopsis* (based on either the 27,873 protein coding and RNA genes, or including the 3,241 novel genes)[2,13], 34% less than rice[3], 46% less than poplar[4] and 19% less than grape[5] (Table 1).

Comparison of the papaya genome with that of *Arabidopsis* sheds new light on angiosperm evolutionary history in several ways. Considering only the 200 longest papaya scaffolds, we found 121 co-linear blocks. The papaya blocks range in size from 1.36 Mb containing 181 genes to 0.16 Mb containing 19 genes (a statistical, rather than a biological, lower limit); the corresponding *Arabidopsis* regions range from 0.69 Mb containing 163 genes to 60 kilobases (kb) containing 18 genes. Across the 121 papaya segments for which co-linearity can be detected, 26 show primary correspondence (that is, excluding the effects of ancient triplication detailed below) to only one *Arabidopsis* segment, 41 to two, 21 to three, 30 to four, and only 3 to more than four.

The fact that many papaya segments show co-linearity with two to four *Arabidopsis* segments (Fig. 1, and Supplementary Figs 3 and 4) is most parsimoniously explained if either one or two genome duplications have affected the *Arabidopsis* lineage since its divergence from papaya. Although it was suspected that the most recent *Arabidopsis* genome duplication, α[14], might affect only a subset of the Brassicales[15], previous phylogenetic dating of these events[15] had suggested that the more ancient β-duplication occurred early in the eudicot radiation, well before the *Arabidopsis*–*Carica* divergence. This incongruity is under investigation.

In contrast, individual *Arabidopsis* genome segments correspond to only one papaya segment, indicating that no genome duplication has occurred in the papaya lineage since its divergence from *Arabidopsis* about 72 million years ago[5]. The lack of relatively recent papaya genome doubling is further supported by an L-shaped distribution of intra-EST correspondence for papaya (not shown). However, multiple genome/subgenome alignments (see Supplementary Methods) reveal evidence in papaya of the ancient 'γ' genome duplication shared with *Arabidopsis* and poplar that is postulated to have occurred near the origin of angiosperms[14]. Indeed, both papaya (with no subsequent duplication) and poplar (with a relatively low rate of duplicate gene loss) suggest that γ was not a duplication but a triplication (Fig. 1), with triplicated patterns evident for about 25% of the 247 Mb comprising the 200 largest papaya scaffolds.

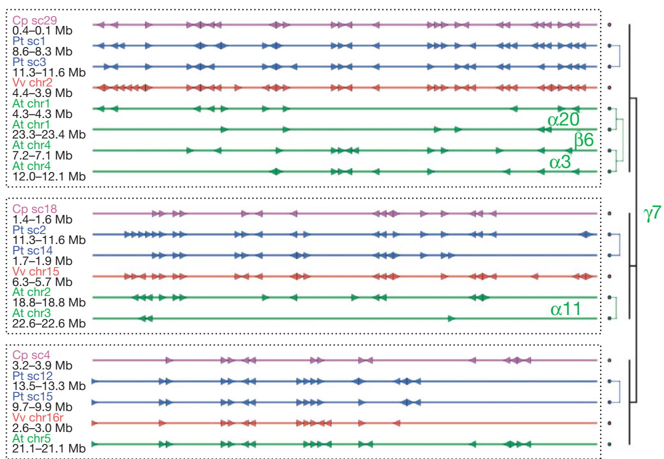



**Figure 1 | Alignment of co-linear regions from *Arabidopsis* (green), papaya (magenta), poplar (blue) and grape (red).** 'Vv chr16r' is an unordered ultracontig that has been assigned to grape chromosome 16. Triangles represent individual genes with transcriptional orientations. Several *Arabidopsis* regions belong to previously identified duplication segments (α3, α11, α20, β6, γ7, shown to the right)[23]. The whole syntenic alignment supports four distinct whole-genome duplication events: α, β within the *Arabidopsis* lineage, an independent duplication in poplar, and γ which is shared by all four eudicot genomes. Co-linear regions can be grouped into three γ sub-genomes based on Camin–Sokal parsimony criteria.

**Table 1 | Statistics of sequenced plant genomes**

| | *Carica papaya* | *Arabidopsis thaliana* | *Populus trichocarpa* | *Oryza sativa* (japonica) | *Vitis vinifera* |
|---|---|---|---|---|---|
| Size (Mbp) | 372 | 125 | 485 | 389 | 487 |
| Number of chromosomes | 9 | 5 | 19 | 12 | 19 |
| G + C content total (%) | 35.3 | 35.0 | 33.3 | 43.0 | 36.2 |
| Gene number | 24,746 | 31,114* | 45,555 | 37,544 | 30,434 |
| Average gene length (bp per gene) | 2,373 | 2,232 | 2,300 | 2,821 | 3,399 |
| Average intron length (bp) | 479 | 165 | 379 | 412 | 213 |
| Transposons (%) | 51.9 | 14 | 42 | 34.8 | 41.4 |

* The gene number of *Arabidopsis* is based on the 27,873 protein-coding and RNA genes from The Arabidopsis Information Resource website (http://www.arabidopsis.org/portals/genAnnotation/genome_snapshot.jsp) and recently published 3,241 novel genes[6].

This is most probably an underestimate that will increase as papaya contiguity is improved. Triplication in papaya and poplar corresponds closely to the triplication suggested by an independent analysis of the grape genome[5].

A few hundred papaya chromosomal segments were aligned using BLASTZ to their one to four syntenic regions in *Arabidopsis*, and the results examined visually using the Genome Evolution (GEvo) viewer[16]. The orthologous region of grape was also included[5], making the alignment a six-way comparison. One example is given in Supplementary Fig. 5: a 500 kb segment of papaya, its four 60 kb syntenic, orthologous *Arabidopsis* segments and the 400 kb orthologous segment of grape.

For the homologous *Arabidopsis* segments that are discernibly co-linear (by MC-SCANNER) to the 200 longest papaya scaffolds, 34.8% of *Arabidopsis* genes in any one segment correspond to a papaya gene, whereas only 24.8% of papaya genes in any one segment correspond to an *Arabidopsis* gene. Moreover, the *Arabidopsis* homologous segments contain fewer genes, on average only about 57.9% of the number in their papaya counterparts.

Papaya provides a useful outgroup necessary to detect subfunctionalization. Supplementary Fig. 6 is a GEvo screenshot of a blastn alignment illustrating subfunctionalization of conserved non-coding sequences (CNSs)[17] upstream of two syntenic, duplicate *Arabidopsis* genes and their single papaya orthologous gene. The α-duplicated genomes within *Arabidopsis* are perfect for CNS discovery[18].

Comparative analysis of the papaya and *Arabidopsis* 5′ untranslated regions showed that only 14% of orthologous promoter pairs exhibit significantly higher levels of sequence identity than random comparisons (Supplementary Figs 7 and 8). Although some highly conserved promoters show substantial conservation across much of their length, sequence similarity for most orthologous papaya promoters is indistinguishable from background.

Global analysis of all inferred protein models from papaya, *Arabidopsis*, poplar, grape and rice clusters the 208,901 non-redundant protein sequences into 39,706 similarity groups, or 'tribes'[19], 11,851 of which contain two or more genes (see Supplementary Methods). Tribes with multiple genes in a species typically correspond to families or subfamilies of genes; however, tribes may also contain just one gene ('singleton tribes'). In papaya, 25,312 gene models were classified into 12,958 tribes, 5,669 of which were specific to papaya (Supplementary Table 4). Of the papaya-specific tribes, 5,314 were singleton tribes. EST support was markedly lower for genes in papaya-specific tribes (below 14%) than in tribes that included genes from at least one other taxon (72.4%).

To investigate the smaller number of genes in papaya, we compared tribe membership from each of the five sequenced angiosperm species (Supplementary Table 5). Among the 6,726 tribes that contain genes from both *Arabidopsis* and papaya, 3,595 contain equal numbers of genes from both species. However, tribes with more *Arabidopsis* genes outnumber those with more papaya genes by more than 2:1 (2,153:979). The trend of smaller number of papaya genes is widespread across tribes of all sizes and major functional categories (Supplementary Table 6 and Supplementary Fig. 9).

We then examined membership in the 815 tribes with members identified as being likely transcription factors in the *Arabidopsis* transcription factor database (http://arabidopsis.med.ohio-state.edu/AtTFDB/). This set includes 2,897 genes in *Arabidopsis* and 2,438 in papaya (a ratio of 1.19:1). The details of tribe membership are illustrated for 25 exemplar families and superfamilies (Fig. 2), where most transcription-factor tribes have fewer genes in papaya than *Arabidopsis*. Some transcription-factor tribes had more genes in papaya, specifically RWP-RK, MADS-box, Scarecrow, TCP and Jumonji gene families. Interestingly, the difference in MADS protein family size appears to be due to expanded numbers for half of the 36 MADS tribes. The other 18 MADS tribes had fewer papaya genes, including 14 that were not found in papaya.

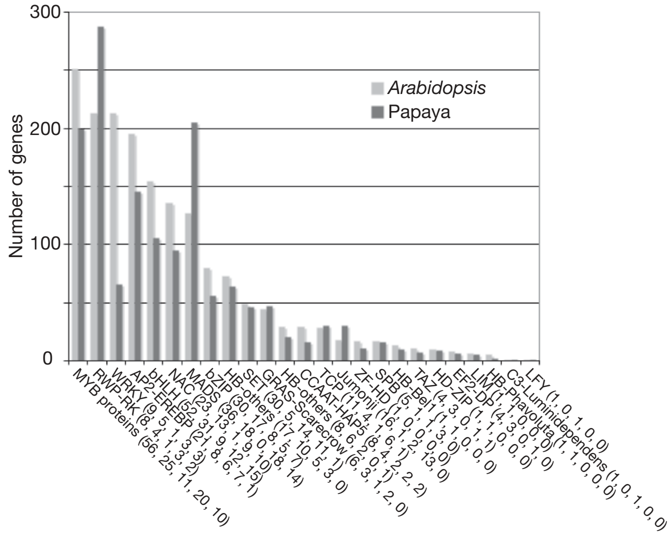



**Figure 2 | Comparison of gene numbers in transcription-factor tribe or related tribes from *Arabidopsis* and papaya.** Most transcription factors are represented by fewer genes in papaya than *Arabidopsis*. Transcription-factor names are given, with values after the names corresponding to: number of tribes with genes assigned to transcription factor group, number of tribes with smaller counts in papaya than *Arabidopsis*, number of tribes with equal counts in papaya and *Arabidopsis*, number of tribes with larger counts in papaya, and number of tribes with zero members in papaya. Supporting data are provided in Supplementary Table 8.

Assuming that a generalized angiosperm could potentially require only the types and minimal numbers of genes that are shared among divergent plant species, we examined each of the tribes shared among the five angiosperms with sequenced genomes. The number of genes required in a minimal flowering plant is based on the observed minimum number of genes across each of the shared tribes (Table 2). When the smallest observed number is taken for each evolutionarily conserved tribe, a minimal angiosperm genome of 13,311 genes is estimated. Papaya has the smallest number of genes for more tribes than any other sequenced taxon (4,515, or 76% of 5,925 shared tribes), reinforcing the notion that papaya has fewer genes than any angiosperm sequenced so far.

Only 55 nucleotide-binding site (NBS)-containing R genes were identified in papaya; about 28% of the 200 NBS genes in *Arabidopsis*[20] and less than 10% of the 600 NBS genes in rice[21]. Resistance proteins also have a carboxy-terminal leucine-rich repeat (LRR) domain. These NBS-containing R-gene families can be subdivided into three classes: NBS–LRR, toll interleukin receptor (TIR)–NBS–LRR, and coiled-coil (CC)–NBS–LRR on the basis of their amino-terminal region. Papaya NBS–LRR outnumbered both TIR–NBS–LRR and CC–NBS–LRR genes, in contrast to both poplar (with more CC–NBS–LRR genes[4]) and *Arabidopsis* (with more TIR–NBS–LRR). More than 50% of the NBS-type R genes were clustered in about eight scaffolds, indicating that resistance gene evolution may involve duplication and divergence of linked gene families.

**Table 2 | Deduced potential minimal angiosperm gene number based on species with smallest number of genes for each tribe**

| | *Carica papaya* | *Arabidopsis thaliana* | *Populus trichocarpa* | *Oryza sativa* (japonica) | *Vitis vinifera* | Shared tribes | Minimal gene number |
|---|---|---|---|---|---|---|---|
| Shared tribes with minimum | 4,515 | 3,597 | 1,548 | 3,657 | 3,597 | 5,925 | 13,331 |
| Number of unique tribes | 5,708 | 2,950 | 6,338 | 13,003 | 3,567 | | |
| Number of conserved tribes lost or missing from each species | 405 | 113 | 28 | 429 | 175 | | |

Homologues for genes involved in cellulose biosynthesis are present in papaya and *Arabidopsis*, with more cellulose synthase genes in poplar, perhaps associated with wood formation. Papaya has at least 32 putative β-glucosyl transferase (GT1) genes compared with 121 in *Arabidopsis* identified using sequence alignment. A total of 38 and 40 cellulose synthase-related genes (GT2) were identified in papaya using the 48 poplar and 31 *Arabidopsis* genes as queries, respectively. These genes include 11 cellulose synthase (CesA) genes, the same number as in *Arabidopsis* but 7 fewer than in poplar. Putative cellulose orientation genes (COBRA) were more abundant in *Arabidopsis* (12) than in papaya (8).

Papaya also has a similar complement though fewer genes for cell-wall synthesis than *Arabidopsis*. Papaya and *Arabidopsis*, respectively, have 6 and 12 callose synthase genes (GT2); 15 and 15 xyloglucan α-1,2-fucosyl transferases (GT37); 5 and 7 β-glucuronic acid transferases in familes GT43 and GT47; and 27 and 42 in GT8 that includes galacturonosyl transferases, associated with pectin synthesis.

The cell wall of plants is capable of both plastic and elastic extension, and controls the rate and direction of cell expansion[22]. Despite fewer whole-genome duplications, papaya has a similar number of putative expansin A genes (24) as *Arabidopsis* (26) and poplar (27), and more expansin B genes (10) than *Arabidopsis* (6) and poplar (3).

In contrast to expansion-related genes, papaya has on average about 25% fewer cell-wall degradation genes than *Arabidopsis*, in some cases far fewer. For example, papaya and *Arabidopsis*, respectively, have 4 and 12 endoxylanase-like genes in glycoside hydrolase family 10 (GH10); 29 and 67 pectin methyl esterases (carbohydrate esterase family 8); 28 and 69 polygalacturonases (GH28); 15 and 49 xyloglucan endotransglycosylase/hydrolases (GH16); 18 and 25 β-1,4-endoglucanases (GH9); 42 and 91 β-1,3-glucanases (GH17); and 15 and 27 pectin lyases (PL1).

A semi-woody giant herb that accumulates lignin in the cell wall at an intermediate level between *Arabidopsis* and poplar, papaya generally has intermediate numbers of lignin synthetic genes, fewer than poplar but more than *Arabidopsis* despite fewer opportunities for duplication in papaya. Poplar, papaya and *Arabidopsis* have 37, 30 and 18 candidate genes for the lignin synthesis pathway, respectively[4,23], with papaya having an intermediate number of genes for the PAL, C4H, 4CL and HCT gene families, and only one COMT and two C3H genes. In contrast, poplar has three C3H genes, which are presumed to convert *p*-coumaroyl quinic acid to caffeoyl shikimic acid, whereas there are two in papaya and one in *Arabidopsis*. Papaya, *Arabidopsis* and poplar each have two genes in the family CCoAOMT, which are presumed to convert caffeic acid to ferulic acid[4]. Compared with these other plants, papaya has the fewest genes in the CCR gene family (1 gene) and the most in the F5H (4 genes) and CAD gene families (18 genes), which all mediate later steps of the lignin biosynthesis pathway.

More starch-associated genes in papaya, a perennial, may be due to a greater need for storage in leaves, stem and developing fruit than in *Arabidopsis*, an ephemeral that stores oil in the seed. Papaya and *Arabidopsis*, respectively, have 13 and 6 putative starch synthase (GT5) genes; 8 and 3 starch branching genes; 6 and 3 isoamylases (GH13); and 12 and 9 β-amylases (GH14). Early unloading of fruit sugar in papaya is probably symplastic[24], with five genes for sucrose synthase/sucrose phosphate synthase (GT4); seven are reported for *Arabidopsis*. Five acid invertase (GH32) sequences were found in papaya whereas 11 have been reported in *Arabidopsis*. Papaya has at least seven putative neutral invertase (GH32) genes; *Arabidopsis* has six. Wall-associated kinases (WAK) are thought to be involved in the regulation of vacuolar invertases, with 17 in *Arabidopsis* and only 10 in papaya. *Arabidopsis* and papaya have 14 and 7 hexose transporters, respectively. The greater number of genes for sugar accumulation in *Arabidopsis* may reflect recent genome duplications.

Papaya has undergone particularly striking amplification of genes involved in volatile development. Papaya and *Arabidopsis*, respectively, have 18 and 8 genes for cinnamyl alcohol dehydrogenase; 2 and 1 genes for cinnamate-4-hydroxylase; 9 and 3 genes for phenylalanine ammonia lyase; and 24 and 3 limonene cyclase genes.

Papaya ripening is climacteric, with the rise in ethylene production occurring at the same time as the respiratory increase[25]. Papaya and *Arabidopsis*, respectively, have similar numbers of genes involved in ethylene synthesis, with four each for *S*-adenosyl methionine synthase (SAM synthase); 8 and 13 for aminocyclopropane carboxylic acid (ACC) synthase (ACS); 8 and 12 for ACC oxidase (ACO); and 42 and 64 for ethylene-responsive binding factors (AP2/ERF).

Because papaya grows in tropical climates where daily light/dark cycles do not change much over the year, we can ask if more or fewer light/circadian genes are required to synchronize with the environment. In fact, there are fewer light/clock genes in the papaya genome (49% and 34% of poplar and *Arabidopsis*, respectively; Supplementary Table 7). However, among the core circadian clock genes, the pseudo-response regulators (PRRs; Supplementary Fig. 10) have expanded in poplar compared with *Arabidopsis*, and the papaya PRR7 cluster has seemingly duplicated with the recent poplar salicoid-specific genome duplication[4] (Supplementary Fig. 11). Against the backdrop of fewer overall genes, the parallel expansion of the PRRs is consistent with circadian timing being important in papaya.

The PAS–FBOX–KELCH genes control light signalling and flowering time; however, the only papaya orthologue (ZTL) lacks an obvious KELCH domain compared with *Arabidopsis* and poplar, which have five and one KELCH domains, respectively (Supplementary Fig. 10). In fact, the papaya genome contains fewer KELCH domains (37 compared with 130 and 74 in *Arabidopsis* and poplar, respectively). In contrast, there are three constitutive photomorphogenic 1 (COP1) paralogues in the papaya genome compared with only one in *Arabidopsis* (Supplementary Tables 7 and 8). A similar expansion has been noted in moss (*Physcomitrella patens*), which has nine COP1 paralogues that are hypothesized to aid in tolerance to ultraviolet light (Supplementary Fig. 12)[26]. Both KELCH domains and the WD-40 of the COP1 family form β-propellers and play a role in light-mediated ubiquitination. There is not a general expansion of WD-40 genes in papaya (173 compared with 227 in *Arabidopsis*). Perhaps papaya has developed an alternative way of integrating light or timing information specific to day-neutral plants, such as a strict adherence to the diel light/dark cycle that is better served by the COP-mediated system.

Sex determination in papaya is controlled by a pair of primitive sex chromosomes, with a small male-specific region of the Y chromosome (MSY)[8]. The physical map of the MSY is currently estimated by chromosome walking to span about 8 Mb (ref. 27). Two scaffolds in the current female-genome sequence align to the X chromosome physical map based on BAC end sequences, spanning 4.5 Mb and including 254 predicted protein-encoding genes, of which 75 (29.5%) have EST support (Supplementary Table 9 and Supplementary Fig. 13). If adjusted for the percentage of unigene validation for other genes (48.0%), the estimated number of genes in the X-specific region would be 156. The average gene density would be one gene per 19.5 kb, lower than the estimated genome average of one gene per 14.3 kb. By contrast, among seven completely sequenced MSY BACs totalling 1.2 Mb, a total of four expressed genes were found on two of the BACs[14,28]. The somewhat lower-than-average gene density in the X-specific scaffolds is accompanied by more repetitive DNA (58.3%) than the genome-wide average, perhaps because this region is near the centromere[28]. Re-analysis of the repetitive DNA content of the MSY BACs, to include the new papaya-specific repeat families identified herein, increased the average repeat sequence to 85.6%, with 54.1% Gypsy and 1.9% Copia retro-elements (Supplementary Table 10). This compares with an earlier estimate of 17.9% using the *Arabidopsis* repeat database alone[28].

The SunUp genome has presented an opportunity to analyse transgene insertion sites critically. Southern blot analysis was key in the initial identification of transgenic insertion fragments and was performed with probes spanning the entire 19,567-bp transformation vector used for bombardment (Supplementary Fig. 14). Among the identified inserts were the functional coat-protein transgene conferring resistance to papaya ringspot virus, which was found in an intact 9,789-bp fragment of the transformation plasmid, and a 1,533-bp fragment composed of a truncated, non-functional *tetA* gene and flanking vector backbone sequence. The structures of the coat-protein transgene and *tetA* region insertion sites were determined from cloned sequences. Southern analysis also confirmed a 290-bp non-functional fragment of the *npt*II gene originally identified by WGS sequence analysis (Supplementary Fig. 15). Five of the six flanking sequences of the three insertions are nuclear DNA copies of papaya chloroplast DNA fragments. The integration of the transgenes into chloroplast DNA-like sequences may be related to the observation that transgenes produced either by *Agrobacterium*-mediated or biolistic transformation are often inserted in AT-rich DNA[29], as is the chloroplast DNA of papaya and other land plants. Four of the six insert junctions have sequences that match topoisomerase I recognition sites, which are associated with breakpoints in genomic DNA transgene insertion sites and transgene rearrangements[29]. The presence of these inserts was confirmed by high-throughput MUMmer[30] analysis for each region of the transformation vector. Evidence for the presence of other transgene inserts is not conclusive (Supplementary Note 3).

Its lower overall gene number notwithstanding, striking variations in gene number within particular functional groups, superimposed on the average approximate 20% reduction in papaya gene number relative to *Arabidopsis*, may be related to key features of papaya morphological evolution. Despite a closer evolutionary relationship to *Arabidopsis*, papaya shares with poplar an increased number of genes associated with cell expansion, consistent with larger plant size; and lignin biosynthesis, consistent with the convergent evolution of tree-like habit. Amplification of starch-synthesis genes in papaya relative to *Arabidopsis* is consistent with a greater need for storage in leaves, stem and developing fruit of this perennial. Tremendous amplification in papaya of genes related to volatile development implies strong natural selection for enhanced attractants that may be key to fruit (seed) dispersal by animals and which may also have attracted the attention of aboriginal peoples. This also foreshadows what we might expect to discover in the genomes of other fragrant-fruited trees, as well as plants with striking fragrance of leaves (herbs), flowers or other organs.

Arguably, the sequencing of the genome of SunUp papaya makes it the best-characterized commercial transgenic crop. Because papaya ringspot virus is widespread in nearly all papaya-growing regions, SunUp could serve as a transgenic germplasm source that could be used to breed suitable cultivars resistant to the virus in various parts of the world. The characterization of the precise transgenic modifications in SunUp papaya should also serve to lower regulatory barriers currently in place in some countries.

## METHODS SUMMARY

**Gene annotation.** Papaya unigenes from complementary DNA were aligned to the unmasked genome assembly, which was then used in training *ab initio* gene prediction software. Spliced alignments of proteins from the plant division of GenBank, and transcripts from related angiosperms, were generated. Gene predictions were combined with spliced alignments of proteins and transcripts to produce a reference gene set. Detailed descriptions are given in Methods.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 6 September 2007; accepted 22 February 2008.**

1. Gonsalves, D. Control of papaya ringspot virus in papaya: a case study. *Annu. Rev. Phytopathol.* **36**, 415–437 (1998).
2. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
3. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
4. Tuskan, G. A. *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604 (2006).
5. Jaillon, C. O. *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
6. Arumuganathan, K. & Earle, E. D. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–218 (1991).
7. Fitch, M. M. M., Manshardt, R. M., Gonsalves, D., Slightom, J. L. & Sanford, J. C. Virus resistant papaya plants derived from tissues bombarded with the coat protein gene of papaya ringspot virus. *Bio/technology* **10**, 1466–1472 (1992).
8. Liu, Z. *et al.* A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427**, 348–352 (2004).
9. Wikström, N., Savolainen, V. & Chase, M. W. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B* **268**, 2211–2220 (2001).
10. Storey, W. B. Papaya. in *Outlines of Perennial Crop Breeding in the Tropics* (eds Ferwerda, F. P. and Wit, F.) 389–408 (H. Veenman & Zonen, Wageningen, 1969).
11. Li, L. *et al.* Genome-wide transcription analyses in rice using tiling microarrays. *Nature Genet.* **38**, 124–129 (2006).
12. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
13. Hanada, K., Zhang, X., Borevitz, J. O., Li, W.-H. & Shiu, S.-H. A large number of novel coding small open reading frames in the intergenic regions of the *Arabidopsis thaliana* genome are transcribed and/or under purifying selection. *Genome Res.* **17**, 632–640 (2007).
14. Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438 (2003).
15. Schranz, M. E. & Mitchell-Olds, T. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* **18**, 1152–1165 (2006).
16. Lyons, E. & Freeling, M. How to usefully compare homologous plant genes and chromosomes as DNA sequence. *Plant J.* **53**, 661–673 (2008).
17. Inada, D. C. *et al.* Conserved noncoding sequences in the grasses. *Genome Res.* **13**, 2030–2041 (2003).
18. Thomas, B. C., Rapaka, L., Lyons, E., Pedersen, B. & Freeling, M. *Arabidopsis* intragenomic conserved noncoding sequence. *Proc. Natl Acad. Sci. USA* **104**, 3348–3353 (2007).
19. Wall, P. K. *et al.* PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res.* **36**, D970–D976 (2008).
20. Meyers, B. C., Morgante, M. & Michelmore, R. W. TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *Plant J.* **32**, 77–92 (2002).
21. Zhou, T. *et al.* Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genomics* **271**, 402–415 (2004).
22. Fry, S. C. Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. *New Phytol.* **161**, 641–675 (2004).
23. Ehlting, J. *et al.* Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant J.* **42**, 618–640 (2005).
24. Zhou, L. L. & Paull, R. E. Sucrose metabolism during papaya (*Carica papaya*) fruit growth and ripening. *J. Am. Soc. Hortic. Sci.* **126**, 351–357 (2001).
25. Paull, R. E. & Chen, N. J. Postharvest variation in cell wall-degrading enzymes of papaya (*Carica papaya* L.) during fruit ripening. *Plant Physiol.* **72**, 382–385 (1983).
26. Richardt, S., Lang, D., Reski, R., Frank, W. & Rensing, S. A. PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. *Plant Physiol.* **143**, 1452–1466 (2007).
27. Yu, Q. *et al.* Low X/Y divergence of four pairs of papaya sex-liked genes. *Plant J.* **53**, 124–132 (2008).
28. Yu, Q. *et al.* Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. *Mol. Genet. Genomics* **278**, 177–185 (2007).
29. Sawasaki, T., Takahashi, M., Goshima, N. & Morikawa, H. Structures of transgene loci in transgenic *Arabidopsis* plants obtained by particle bombardment: junction regions can bind to nuclear matrices. *Gene* **218**, 27–35 (1998).
30. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank X. Wan, J. Saito and A. Young at the University of Hawaii for technical assistance; C. Detter at the DOE Joint Genome Institute; F. MacKenzie, O. Veatch and T. Uhm at the Hawaii Agriculture Research Center; L. Li, W. Teng, Y. Wu, Y. Yang, C. Zhou, N. Wang, P. Wang and D. Fei at the Tianjin Biochip Corporation, Tianjin Economic-Technological Development Area, Tianjin; and R. Herdes, L. Diebold, R. Kim, A. Hernandez, S. Ali and L. Bynum at the University of Illinois at Urbana-Champaign. This papaya genome-sequencing project was given support by the University of Hawaii and the US Department of Defense grant number W81XWH0520013 to M.A., the Maui High Performance

Computing Center to M.A., the Hawaii Agriculture Research Center to R.M. and Q.Y., and Nankai University, China, to L.W. Other support to the papaya genome project included the United States Department of Agriculture T-STAR program; a United States Department of Agriculture–Agricultural Research Service cooperative agreement (CA 58-3020-8-134) with the Hawaii Agriculture Research Center; the University of Illinois; the National Science Foundation Plant Genome Research Program; and Tianjin Municipal Special Fund for Science and Technology Innovation Grant 05FZZDSH00800. We thank P. Englert, former chancellor of the University of Hawaii, for initial infrastructure support of the research.

**Author Information** The papaya WGS sequence is deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under accession number ABIM00000000. The version described in this paper is the first version, ABIM01000000. The GenBank accession numbers of the papaya ESTs are EX227656–EX303501.  Correspondence and requests for materials should be addressed to M.A. (alam@hawaii.edu) or L.W. (wanglei@nankai.edu.cn).

## METHODS

**Genome assembly.** The Genome sequence was assembled by Arachne[31]. WGS reads and BAC end reads were trimmed by LUCY and screened for organellar sequences[32]. Two approaches were applied to screening and removing reads of presumably organellar origin to alleviate the load in assembling highly repetitive regions by WGS assembly software. The first approach was an iterative process, in which reads were assembled, contigs matching with organellar genomes identified, constituent reads removed, and the process repeated by two or three more rounds. This approach produced the read sets for the released assemblies Stripped3 and Stripped4. The second approach was to remove plasmid clones and BAC clones of presumably organellar origin by identifying clones with both end reads matching entirely with organellar genomes, with physical map information an amendment to the identification of BAC clones. Two rounds of iterative screening based on pairing information of assembled and unplaced reads were added to the second approach to generate the read set for the released Papaya1.0 assembly.

The sequence error rates were estimated by aligning assembled shotgun sequences with two finished BACs (GenBank accession numbers EF661023 and EF661026). The error rate of the assembly at 3× coverage or deeper (74.2% of assembled sequences) was less than 0.01% based on average quality values of 20 or greater in trimmed sequence. The error rate at 2× coverage (16.3%) was 0.37%. The error rate at 1× coverage (9.5%) was approximately 0.75%, because these sequences are at the ends of the contigs (and sequence reads) where the sequence quality declined.

**Genome annotation.** Gene annotation was conducted following the TIGR Eukaryotic Annotation Pipeline. Repeat sequences were identified in the assembled genome and masked by RepeatMasker, RepeatScout and TransposonPSI, based on known repeat elements in RepBase databases and TIGR Plant Repeat Databases, and the papaya novel repeat database constructed in this study[33,34]. Program to Assemble Spliced Alignments (PASA)[35] was used to generate spliced alignments of papaya unigenes to the unmasked assembly, which was then used in training *ab initio* gene prediction software Augustus, GlimmerHMM and SNAP[36–38]. *Ab initio* gene prediction software Fgenesh, Genscan and TWINSCAN were trained on *Arabidopsis*[39–41]. Spliced alignments of proteins from the plant division of GenBank and transcripts from related angiosperms (*Arabidopsis thaliana*, *Glycine max*, *Gossypium hirsutum*, *Medicago truncatula*, *Nicotiana tabacum*, *Oryza sativa*, *Zea mays*) were generated by the Analysis and Annotation Tool (AAT)[42]. Spliced alignment of proteins from the Pfam database were generated using GeneWise[43,44]. Gene predictions generated by Augustus, Fgenesh, Genscan, GlimmerHMM, SNAP and TWINSCAN were combined with spliced alignments of proteins and transcripts to produce a reference gene set using the evidence-based combiner EVidenceModeler (EVM)[45]. Protein domains were predicted using InterProScan against protein databases (PRINTS, Pfam, ProDom, PROSITE, SMART)[46–50].

**Construction of papaya repeat database.** We used a combination of homology-based and *de novo* methods to identify signatures of transposable elements in the papaya genome. We used RepeatMasker (http://www.repeatmasker.org) in combination with a custom-built library of plant repeat elements for our initial classification of transposable elements. The customized library was generated by combining plant repeats from Repbase and plant repeat databases from TIGR (ftp://ftp.tigr.org/pub/data/TIGR_Plant_Repeats)[33]. Repeat elements identified as ribosomal RNA sequences in the TIGR databases match a large fraction of the papaya genome (about 3%). Ribosomal RNAs were identified separately, and therefore were excluded from our repeat library, leaving a database of 76,924 repeat sequences that were used to search the papaya genome.

Homology-based methods are limited to finding elements that have not diverged too greatly from known repeats. Because databases of known transposable elements are necessarily incomplete, we used additional *de novo* methods to search for repeat elements in papaya contigs. For this, we applied two recently developed repeat-finding tools, PILER and RepeatScout to the complete set of contigs from the papaya genome[34,51]. PILER was able to find 428 repeat families whereas RepeatScout found 6,596 repeat sequences.

The repeat families obtained from PILER and RepeatScout were annotated using a combination of manual curation (786 repeat families) and automated analysis. For the automated annotation, the combined data set from PILER and RepeatScout was made non-redundant (using CD-HIT at the 90% similarity level), leaving behind 6,240 repeat families[52]. As a post-processing step, we selected only those families that had at least ten good ($E$ value $< 1 \times 10^{20}$) BLAST matches to papaya contigs. The resulting data set contained 2,198 repeat families in the papaya genome. BLAST searches against non-redundant and PTREP (http://wheat.pw.usda.gov/ITMI/Repeats) were then used to identify repeat families matching genes associated with transposons and retrotransposons. This procedure discovered an additional 103 repeat families that could be annotated as being retrotransposons. The combined database of 889 annotated papaya-specific transposable-element sequences was used in addition to the database of known repeats to annotate the papaya genome. The remaining, unannotated repeat families (1,455 sequences with no matches to known genes) were then used to estimate the additional repeat content of the genome.

31. Jaffe, D. B. *et al.* Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13,** 91–96 (2003).
32. Chou, H. H. & Holmes, M. H. DNA sequence quality trimming and vector removal. *Bioinformatics* **17,** 1093–1104 (2001).
33. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker (Release Open-3.1.3, 2006).
34. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21** (suppl.), i351–i358 (2005).
35. Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31,** 5654–5666 (2003).
36. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (suppl.), ii215–ii225 (2003).
37. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20,** 2878–2879 (2004).
38. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5,** 59 (2004).
39. Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10,** 516–522 (2000).
40. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268,** 78–94 (1997).
41. Korf, I., Flicek, P., Duan, D. & Brent, M. R. Integrating genomic homology into gene structure prediction. *Bioinformatics* **17** (suppl. 1), S140–S148 (2001).
42. Huang, X., Adams, M. D., Zhou, H. & Kerlavage, A. R. A tool for analyzing and annotating genomic sequences. *Genomics* **46,** 37–45 (1997).
43. Finn, R. D. *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res.* **34** (Database issue), D247–D251 (2006).
44. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14,** 988–995 (2004).
45. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9,** R7.1–R7.19 (2008).
46. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33,** W116–W120 (2005).
47. Attwood, T. K. *et al.* PRINTS and its automatic supplement, prePRINTs. *Nucleic Acids Res.* **31,** 400–402 (2003).
48. Bru, C. *et al.* The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Res.* **33** (Database issue), D212–D215 (2005).
49. Hulo, N. *et al.* The PROSITE database. *Nucleic Acids Res.* **34** (Database issue), D227–D230 (2006).
50. Letunic, I. *et al.* SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* **34** (Database issue), D257–D260 (2006).
51. Edgar, R. C. & Myers, E. W. PILER: Identification and classification of genomic repeats. *Bioinformatics* **21** (suppl.), i152–i158 (2005).
52. Li, W. & Godzik, A. CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22,** i1658–i1659 (2006).

LETTERS

# Genetic variation in human *NPY* expression affects stress response and emotion

Zhifeng Zhou[1]*, Guanshan Zhu[1]*†, Ahmad R. Hariri[2], Mary-Anne Enoch[1], David Scott[3], Rajita Sinha[4], Matti Virkkunen[5], Deborah C. Mash[6], Robert H. Lipsky[1], Xian-Zhang Hu[1], Colin A. Hodgkinson[1], Ke Xu[1], Beata Buzas[1], Qiaoping Yuan[1], Pei-Hong Shen[1], Robert E. Ferrell[2], Stephen B. Manuck[2], Sarah M. Brown[2], Richard L. Hauger[7], Christian S. Stohler[8], Jon-Kar Zubieta[3] & David Goldman[1]

**Understanding inter-individual differences in stress response requires the explanation of genetic influences at multiple phenotypic levels, including complex behaviours and the metabolic responses of brain regions to emotional stimuli. Neuropeptide Y (NPY) is anxiolytic[1,2] and its release is induced by stress[3]. NPY is abundantly expressed in regions of the limbic system that are implicated in arousal and in the assignment of emotional valences to stimuli and memories[4–6]. Here we show that haplotype-driven *NPY* expression predicts brain responses to emotional and stress challenges and also inversely correlates with trait anxiety. *NPY* haplotypes predicted levels of *NPY* messenger RNA in post-mortem brain and lymphoblasts, and levels of plasma NPY. Lower haplotype-driven *NPY* expression predicted higher emotion-induced activation of the amygdala, as well as diminished resiliency as assessed by pain/stress-induced activations of endogenous opioid neurotransmission in various brain regions. A single nucleotide polymorphism (SNP rs16147) located in the promoter region alters *NPY* expression *in vitro* and seems to account for more than half of the variation in expression *in vivo*. These convergent findings are consistent with the function of NPY as an anxiolytic peptide and help to explain inter-individual variation in resiliency to stress, a risk factor for many diseases.**

Variation in stress resiliency influences many human characteristics, including both normal and pathological behaviour[7]. Maladaptive responses to stress are critical in the development of many psychiatric disorders, including mood and anxiety disorders[8,9]. Anxiety and emotionality (neuroticism) are moderately to highly heritable traits (40–60%) but are also strongly influenced by exposures to stress in a pattern consistent with gene–environment interaction[10]. These observations point to the importance of genes that modulate the effects of stress. Genes—such as the serotonin transporter—that have so far been implicated in emotional responses have small effects on complex behavioural traits[11], but larger effects on the metabolic responses of the brain to emotional stimuli accessed by brain imaging[12–14].

We evaluated effects of neuropeptide NPY on emotion and stress resiliency using a haplotype-based approach intended to capture effects of unknown loci or locus combinations. We analysed functionally grouped *NPY* haplotypes against a complex behaviour, trait anxiety, and also on intermediate phenotypes accessed by two different brain imaging modalities in which gene effects might be more strongly manifested. An *NPY* seven-marker panel (Supplementary Fig. 1a) genotyped in 516 Finnish Caucasians captured the major haplotypes and linkage disequilibrium features observed in the International HapMap Project (http://www.hapmap.org). A block of strong pairwise linkage disequilibrium encompasses 70% of the gene and extends from the 5′ region to exon 3 (Supplementary Fig. 1b). Five haplotypes (H1–H5) account for 93.8% of chromosomes in this block (Fig. 1a).

We observed haplotype-driven *NPY* mRNA expression in post-mortem brain (US Caucasians, Miami sample) by detecting the differential expression of alleles at single nucleotide polymorphism (SNP) rs5574 C/T, selected because of its high frequency and location in the transcript. Of these 28 samples, chosen because all were heterozygous for rs5574, 16 (57%) showed differential allele expression at an allele ratio of more than 1.2, in either direction. H1 and H4 were low-expression haplotypes, H2 was high, H3 was intermediate and H5 was unclassified because only two H1/H5 heterozygous brains were available (Fig. 1b). This expression-based functional classification is consistent with a cladistically based clustering of haplotypes, indicating that expression variation is linked to gene ancestry (Fig. 1a). The effects on expression of the more common H1, H2 and H3 haplotypes were verified in 47 lymphoblastoid cell lines derived from healthy Finnish men (Fig. 1c) representing the six common diplotypes (72% of all diplotypes). On the basis of lymphoblast *NPY* mRNA levels, the expression value for each haplotype was calculated by regression analysis. Expression values for the six common diplotypes were well predicted under a co-dominant model and had a threefold range (see Supplementary Fig. 2 for details). Diplotypes were clustered into three expression groups: low (LL: H1/H1), intermediate (LH: H1/H3, H3/H3 and H1/H2) and high (HH: H2/H3 and H2/H2) (Fig. 1c). We applied this grouping in subsequent analyses. Two loci, rs3037354 and rs16147, which differentiate the three common haplotypes (see Fig. 1a), correlated with expression (Supplementary Fig. 3). However, *NPY* haplotypes accounted for more variation. Finally, haplotype-predicted *NPY* expression was correlated with plasma NPY peptide levels available in a US sample (New Haven) in both controls ($n = 24$) and alcoholic patients ($n = 18$) sampled during a no-stress condition. Individuals with the low-expression LL diplotype had lower NPY levels than those with high-expression HH diplotypes; individuals with LH diplotypes were intermediate ($P < 0.0001$ in controls and $P = 0.0074$ in alcoholic patients; Fig. 1d).

[1]Laboratory of Neurogenetics, NIAAA, NIH, Bethesda, Maryland 20892, USA. [2]Departments of Psychiatry, Human Genetics, and Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA. [3]Departments of Psychiatry and Radiology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [4]Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06510, USA. [5]Department of Psychiatry, University of Helsinki, 00014, Helsinki, Finland. [6]Department of Neurology, University of Miami School of Medicine, Miami, Florida 33124, USA. [7]Department of Psychiatry, San Diego VA Healthcare System and University of California, San Diego, California 92161, USA. [8]School of Dentistry, University of Maryland, Baltimore, Maryland 21201, USA. †Present address: Innovation Centre China, AstraZeneca Global R&D, Shanghai 201203, China.
*These authors contributed equally to this work.

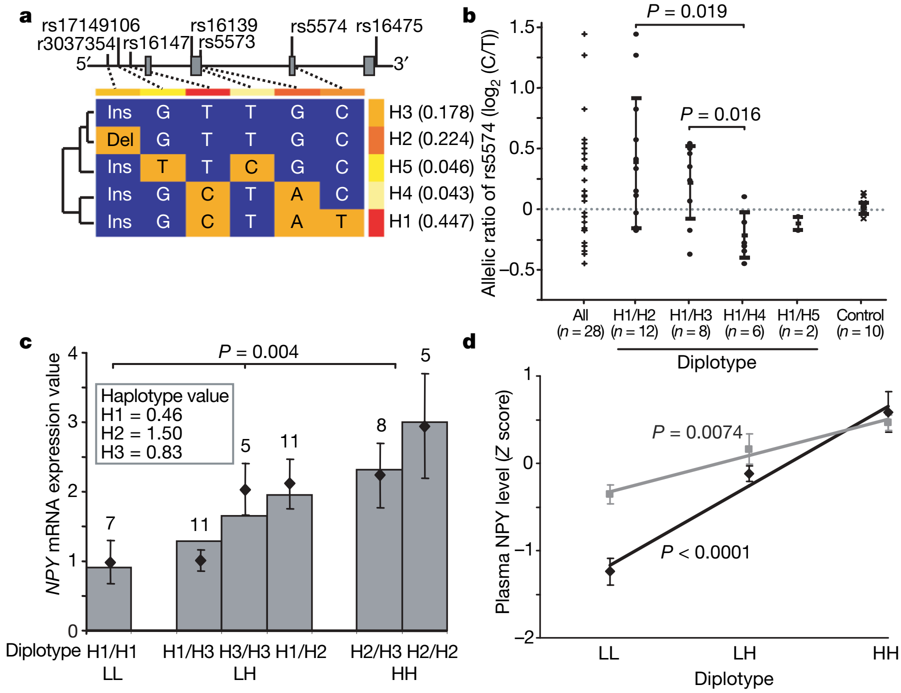



**Figure 1 | Haplotype-predicted *NPY* expression in brain, lymphoblasts and plasma.** **a**, Configuration, frequencies (right) and cladistic clustering (left) of major *NPY* haplotypes (H1, H2, H3, H4 and H5; frequency more than 1%) in Finnish Caucasians ($n = 516$). **b**, Differential allele expression of *NPY* mRNA showing allelic ratio ($\log_2$(C/T)) of rs5574 in 28 heterozygous human post-mortem cerebella (Miami sample) grouped according to diplotypes (columns 2–5, means ± s.d.) to infer expression levels of H2, H3, H4 and H5 relative to H1 (the T allele is found only within H1, as shown in **a**). Diplotypes are compared using a two-tailed *t*-test. Column 1 presents all samples; column 6 shows controls for allelic amplification efficiency. **c**, Expression levels of the six common diplotypes in 47 lymphoblastoid cell lines derived from healthy Finnish Caucasians. Expression values of the three major haplotypes (H1, H2 and H3) were calculated from observed diplotype values (points and error bars show means ± s.e.m.) by multiple regression and were used to predict diplotype values (bars) under a co-dominant model (see Supplementary Fig. 2 for details, including fit to model). Diplotypes were grouped as low (LL), intermediate (LH) or high (HH) in expression. Numbers above columns are *n*. **d**, Diplotype-predicted plasma NPY levels (means ± s.e.m., standardized with *Z* scores; New Haven sample) in controls (black diamonds; LL, $n = 16$; LH, $n = 113$; HH, $n = 39$) and alcoholic patients (grey squares; LL, $n = 29$; LH, $n = 72$; HH, $n = 27$). *P* values were calculated by regression analysis.

These common, functionally significant *NPY* haplotypes were evaluated for their effect on brain responses to emotion and stress. Amygdala activation in response to threat-related facial expressions and other provocative stimuli predicts affective arousal, including anxiety responses[15,16]. We employed a widely used functional magnetic resonance (fMRI) probe[12,17] to assess whether *NPY* diplotypes predicted amygdala reactivity to threat-related facial expressions in 71 healthy volunteers (Pittsburgh sample). This model has been used to identify greater amygdala reactivity in individuals possessing the lower-transcribing allele of the serotonin-transporter-linked polymorphic region (5-HTTLPR)[12,18]. As shown in Fig. 2, amygdala activation in individuals with the low-*NPY*-expression (LL) diplotype was higher than in those with the high-expression (HH) diplotype ($P = 0.003$). *NPY* diplotype predicted amygdala reactivity in an allele-dosage fashion, and it accounted for 9% of the variance in the fMRI amygdala response to emotional challenge. Task-related hippocampal activation was similarly predicted in an allele-dosage fashion ($P = 0.006$; Fig. 2). Functional interactions of the amygdala and hippocampus are crucial for emotional memories, and long-lasting changes in hippocampal architecture are induced by stress[19].

We also tested the ability of *NPY* haplotypes in a model of physical and emotional stress involving moderate levels of sustained muscular pain. This physical and emotional stress activates endogenous opioid neurotransmission in regions of the brain that regulate pain, stress and emotion[20,21]. Endogenous opioid release suppresses pain, stress and anxiety-like responses in animal models[22,23] and humans[21]. The behavioural effects of NPY are mediated, at least in part, through interactions with the endogenous opioid system[24,25]. We measured endogenous opioid release by decreases in the availability of μ-opioid receptors *in vivo* during the painful stressor, quantified by means of positron emission tomography (PET) with the selective μ-opioid receptor radiotracer [$^{11}$C]carfentanil[20,21] (see Supplementary Methods for details). In 35 healthy volunteers (Ann Arbor sample), we found that highly expressed *NPY* diplotypes predicted significantly higher levels of stress-induced μ-opioid system activation in several brain regions (Fig. 3 and Supplementary Table 1) including prefrontal cortex, posterior insula, medial and lateral thalamus, ventral basal ganglia (ventral caudate, ventral putamen and nucleus accumbens) and amygdala (analysis of variance (ANOVA), $P < 0.05$ after correction for multiple comparisons). *NPY* diplotype accounted for 13% of the variance in activation of μ-opioid neurotransmission in the amygdala, 18–35% of the variance in prefrontal cortex, thalamus and nucleus accumbens, and 37% of the variance in posterior insular cortex. In comparison with its effects on the activation of endogenous opioid neurotransmission by painful stress, *NPY* explained less of the variance in the more complex, self-rated pain and affective response phenotypes. *NPY* diplotypes accounted for 3% of the variance in subjective pain (McGill Pain Questionnaire sensory subscale) and 5% of the variance in emotional experience (Positive and Negative Affectivity Scale negative affect) (Supplementary Fig. 4).

In comparison with gene-influenced brain imaging responses in which allele action has been evident even in small data sets[12–14], trait anxiety is a complex behaviour for which gene effects are small[11], and it is the type of gene-influenced behaviour that is perturbed by external factors such as exposure to stress[10,26]. However, it is important to understand the role of *NPY* in complex behaviours and in different contexts. In a relatively modest sample of 137 healthy Finnish Caucasian controls, expression predicted by *NPY* diplotype was inversely correlated with trait anxiety (Fig. 4a), measured with

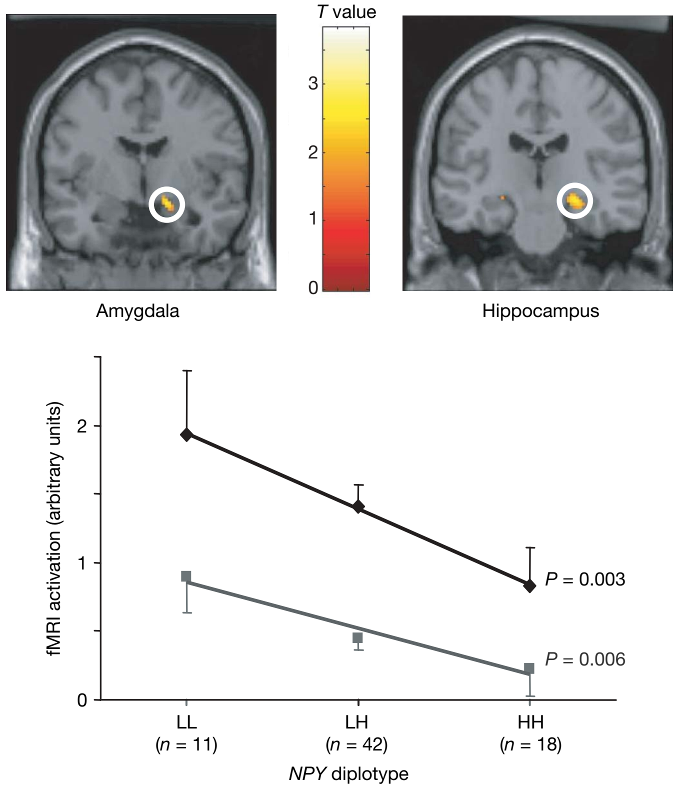



**Figure 2 | Effect of diplotype-predicted *NPY* mRNA expression on fMRI-measured amygdala and hippocampal activation in response to threat-related facial expressions.** Top: statistical parametric maps representing *NPY* diplotype-biased (LL > LH > HH) mean right dorsal amygdala ($x = 16$, $y = -8$, $z = -14$, 64 voxels, $t = 2.82$, $P = 0.003$) and right hippocampal ($x = 24$, $y = -20$, $z = -12$, 132 voxels, $t = 2.56$, $P = 0.006$) activation is shown overlaid on an average sagittal and coronal MRI. Bottom: right amygdala (black diamonds) and hippocampal (grey squares) activities (means and s.e.m.) from clusters grouped by *NPY* diplotypes.

the Tridimensional Personality Questionnaire (TPQ) Harm Avoidance subscales HA1 (Fear of Uncertainty; $P = 0.035$) and HA2 (Anticipatory Worry; $P = 0.031$). There was no correlation with the HA3 (Shyness with Strangers) or HA4 (Fatigability and Asthenia) subscales. Diplotype-predicted *NPY* mRNA expression was lower in the small number of Finnish participants with clinical anxiety disorders ($n = 18$) diagnosed with the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, 3rd edition revised (SCID), in comparison with the same 137 healthy Finnish controls (Fig. 4b). Within Finnish participants with SCID-diagnosed alcoholism ($n = 138$), drug addiction ($n = 38$), anxiety disorders ($n = 18$) and major depression ($n = 22$), we observed no correlation of *NPY*-diplotype-predicted expression with any of the HA subscale scores. This may indicate dysregulation of the stress axis in these patients; in addition, the sample sizes were not large. None of the seven individual markers (Supplementary Fig. 1a) were associated with HA subscales. In contrast with the larger effects of *NPY* diplotype on brain functional responses, the effects of *NPY* on trait anxiety were modest, accounting for 3.3% of the variance in HA1 and 3.4% of the variance in HA2. Very large study samples are required for the consistent detection of gene effects on a complex behaviour such as trait anxiety. We have presented converging results in which modest effects of genes at the level of complex behaviour are supported, and mechanistically clarified, by larger effects on brain-imaging phenotypes reflecting response to emotion and stress.

Population stratification, a potential confounding factor, was addressed with the use of ancestry informative markers (AIMs). As detailed (Supplementary Fig. 5), 186 highly informative AIMs were genotyped; this process was followed by a factor analysis anchored against a panel of 1,017 Centre d'Étude du Polymorphisme Humain (CEPH) worldwide diversity samples representing 52 populations, and yielding ethnic factor scores for each individual. These 186 AIMs yielded a similar seven-factor solution to that observed[27] on the basis of short tandem repeat markers genotyped in the same populations. Our analyses of PET pain/stress response, plasma NPY levels and TPQ Harm Avoidance subscales were not confounded by ethnicity, as revealed by comparisons of individuals above and below medians. For the emotional fMRI imaging sample we had available a set of 15

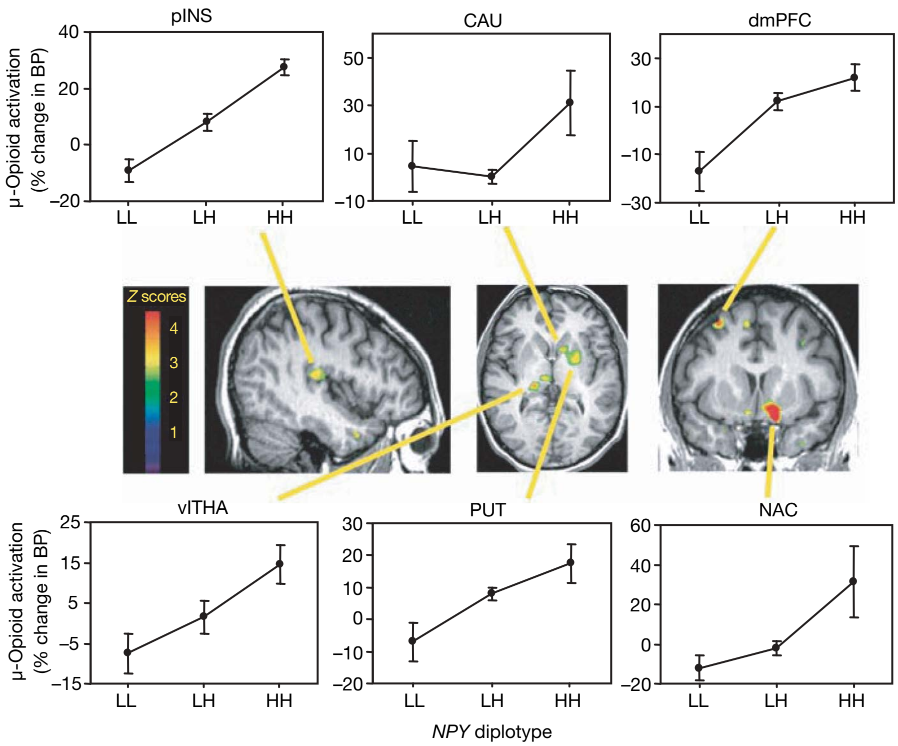



**Figure 3 | Effect of diplotype-predicted *NPY* mRNA expression on pain/stress-induced μ-opioid system activation.** Activation (means ± s.e.m.) measured by the percentage change in the μ-opioid receptor binding potential (BP) among the *NPY* diplotypes (LL, $n = 8$; LH, $n = 21$; HH, $n = 6$) are shown in the dorsomedial prefrontal (dmPFC) and posterior insular (pINS) cortices, ventrolateral thalamus (vlTHA), ventral putamen (PUT) and caudate (CAU) nuclei, nucleus accumbens (NAC) and ventral amygdala (AMY; Supplementary Table 1). *Z* scores represented by the pseudocolour scale (left) are superimposed on an anatomically standardized MRI image. See also Supplementary Table 1 for details of the localization and statistics.

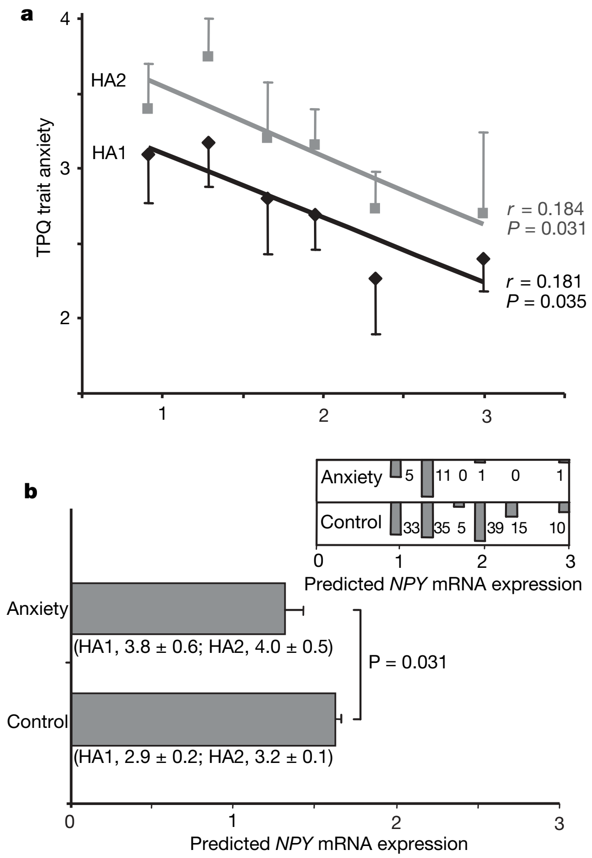



**Figure 4 | Correlations of diplotype-predicted *NPY* mRNA expression with TPQ trait anxiety and anxiety disorders in Finnish Caucasians. a**, Bivariate regression of HA1 and HA2 scores (means and s.e.m.) and diplotype-predicted expression values in 137 healthy subjects (left to right: H1/H1, $n = 33$; H1/H3, $n = 35$; H3/H3, $n = 5$; H1/H2, $n = 39$; H2/H3, $n = 15$; H2/H2, $n = 10$). **b**, Diplotype-predicted *NPY* mRNA levels (means and s.e.m.) of $n = 137$ healthy controls and $n = 18$ anxiety patients by a two-tailed $t$-test (means ± s.e.m. for HA1 and HA2 are also shown). At the top right is a histogram comparing the numbers of controls and anxiety patients within each of the six diplotypes.

AIMs, again with no difference between low and high responders (data not shown).

We identified a locus accounting for part of the *NPY* haplotype effect by testing the four moderately common variants found in the *NPY* promoter region (−1016 base pairs (bp) to 63 bp; see Supplementary Fig. 6a for details) for their ability to influence mRNA expression. The five naturally occurring allele combinations were inserted into a promoterless reporter gene vector (pDsRed2-1) and *NPY* promoter haplotype-driven expression was analysed by transient transfection into the raphe neuronal cell line RN46A. As shown in Supplementary Fig. 6b and Supplementary Table 2, the −399C allele (rs16147) accounted for a 30% decrease in basal expression determined by comparing the allelic variations between the promoter haplotypes. In addition, the TGins allele located at −883 bp (rs3037354) may decrease expression, although to an extent not reaching statistical significance. The combination of −399C and −883TGins reduced expression 47%. These results are consistent with the effect of −399C and the smaller effect of −883TGins on mRNA expression in post-mortem brain and lymphoblastoid cell lines, and with the greater predictive value of haplotypes both *in vitro* and *in vivo* (Fig. 1b, c, and Supplementary Fig. 3), and underlies our choice to emphasize haplotype effects of *NPY*.

Haplotype-based association analysis maximizes the ability to capture information, haplotypes serving as proxies for unknown alleles. However, the existence of multiple haplotypes can lead to a loss of analytic power without some mode of clustering, for example on a cladistic basis[28]. In this study we functionally grouped most *NPY* haplotypes and diplotypes according to levels of expression *in vivo*. There were two less common haplotypes (H4, with a frequency of 4.3%, and H5, with a frequency of 4.6%) that were excluded from association analyses because of a lack of definitive *in vivo* expression data. However, on the basis of *in vitro* data for both (Supplementary Fig. 6b), and limited brain expression data for H4 (Fig. 1b), it is likely that H4 is a low-expression haplotype and H5 is a high-expression haplotype. In addition, the H5 haplotype uniquely contains the missense variant Leu 7→Pro (rs16139C). Several studies have associated Pro 7 with disordered glucose and lipid metabolism[29,30] but not with anxiety, and Pro 7 was not associated with anxiety in our data set (data not shown).

We observed effects of haplotype-predicted *NPY* expression on human trait anxiety and on neurobiological circuits and neurotransmitter systems implicated in the regulation of emotional and stress responses. Although the effect of haplotype-predicted *NPY* expression was modest for trait anxiety, consistent effects were observed across related measures and were more evident for brain metabolic responses to emotional images as measured with fMRI and molecular imaging measures of the activation of the endogenous opioid system after a stressful challenge. These findings indicate the important role of *NPY* in modulating inter-individual variation in emotion and stress resiliency, and reflect the value of a multilevel approach to the genetic analysis of behaviour.

## METHODS SUMMARY

Participants from five independent samples were analysed. A sample of 516 Finnish Caucasians was used to identify *NPY* haplotype configuration and linkage disequilibrium structure, and subsamples were analysed for lymphoblast *NPY* mRNA, trait anxiety, and anxiety disorders. The New Haven sample ($n = 42$) was measured for plasma NPY. The Pittsburgh sample ($n = 71$) received fMRI after an emotional probe. The Ann Arbor sample ($n = 35$) underwent [$^{11}$C]carfentanil PET during a pain/stress challenge. A collection of post-mortem brains from the Miami sample was genotyped and 28 heterozygous cerebella were used for differential allele expression. A full description of participants' demographic data and psychiatric diagnoses is provided in Supplementary Methods. Informed consent was obtained under the auspices of human research protocols approved by institutional review boards of the National Institutes of Health, the University of Helsinki, the University of Pittsburgh, the University of Michigan, and Yale University. *NPY* genotyping was performed with 5′-nuclease assays. Genotyping with 186 AIMs was performed with an Illumina 1536-SNP array. *NPY* mRNA quantification in lymphoblastoid cell lines was performed with reverse transcription (RT) followed by TaqMan real-time PCR. Detection of allele-specific *NPY* expression in post-mortem brain was performed with an RT-coupled 5′ nuclease assay. Plasma concentrations of NPY were measured by radioimmunoassay. DNAs containing *NPY* promoter haplotypes were inserted into a reporter vector, pDsRed2-1, and transfected to the RN46A cell line for the analysis of promoter variants. Details of the fMRI emotional model, the blood-oxygenation-level-dependent (BOLD) fMRI acquisition parameters, the experimental design for pain stressor challenge, the PET scan processes, image reconstruction, and the data analysis are provided in Supplementary Methods.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 13 November 2007; accepted 20 February 2008.**
**Published online 2 April 2008.**

1. Broqua, P., Wettstein, J. G., Rocher, M. N., Gauthier-Martin, B. & Junien, J. L. Behavioral effects of neuropeptide Y receptor agonists in the elevated plus-maze and fear-potentiated startle procedures. *Behav. Pharmacol.* **6**, 215–222 (1995).
2. Heilig, M., Soderpalm, B., Engel, J. A. & Widerlov, E. Centrally administered neuropeptide Y (NPY) produces anxiolytic-like effects in animal anxiety models. *Psychopharmacology (Berl.)* **98**, 524–529 (1989).
3. Thorsell, A., Carlsson, K., Ekman, R. & Heilig, M. Behavioral and endocrine adaptation, and up-regulation of NPY expression in rat amygdala following repeated restraint stress. *Neuroreport* **10**, 3003–3007 (1999).
4. Adrian, T. E. *et al.* Neuropeptide Y distribution in human brain. *Nature* **306**, 584–586 (1983).

5. Allen, Y. S. *et al.* Neuropeptide Y distribution in the rat brain. *Science* **221,** 877–879 (1983).
6. Tatemoto, K., Carlquist, M. & Mutt, V. Neuropeptide Y—a novel brain peptide with structural similarities to peptide YY and pancreatic polypeptide. *Nature* **296,** 659–660 (1982).
7. McEwen, B. S. & Stellar, E. Stress and the individual. Mechanisms leading to disease. *Arch. Intern. Med.* **153,** 2093–2101 (1993).
8. McEwen, B. S. Allostasis and allostatic load: implications for neuropsychopharmacology. *Neuropsychopharmacology* **22,** 108–124 (2000).
9. Sinha, R., Garcia, M., Paliwal, P., Kreek, M. J. & Rounsaville, B. J. Stress-induced cocaine craving and hypothalamic–pituitary–adrenal responses are predictive of cocaine relapse outcomes. *Arch. Gen. Psychiatry* **63,** 324–331 (2006).
10. Caspi, A. *et al.* Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* **301,** 386–389 (2003).
11. Sen, S., Burmeister, M. & Ghosh, D. Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **127,** 85–89 (2004).
12. Hariri, A. R. *et al.* Serotonin transporter genetic variation and the response of the human amygdala. *Science* **297,** 400–403 (2002).
13. Heinz, A. *et al.* Amygdala–prefrontal coupling depends on a genetic variation of the serotonin transporter. *Nature Neurosci.* **8,** 20–21 (2005).
14. Pezawas, L. *et al.* 5-HTTLPR polymorphism impacts human cingulate–amygdala interactions: a genetic susceptibility mechanism for depression. *Nature Neurosci.* **8,** 828–834 (2005).
15. Haas, B. W., Omura, K., Constable, R. T. & Canli, T. Emotional conflict and neuroticism: personality-dependent activation in the amygdala and subgenual anterior cingulate. *Behav. Neurosci.* **121,** 249–256 (2007).
16. Somerville, L. H., Kim, H., Johnstone, T., Alexander, A. L. & Whalen, P. J. Human amygdala responses during presentation of happy and neutral faces: correlations with state anxiety. *Biol. Psychiatry* **55,** 897–903 (2004).
17. Tessitore, A. *et al.* Dopamine modulates the response of the human amygdala: a study in Parkinson's disease. *J. Neurosci.* **22,** 9099–9103 (2002).
18. Hariri, A. R. *et al.* A susceptibility gene for affective disorders and the response of the human amygdala. *Arch. Gen. Psychiatry* **62,** 146–152 (2005).
19. McEwen, B. S. Stress and hippocampal plasticity. *Annu. Rev. Neurosci.* **22,** 105–122 (1999).
20. Zubieta, J. K. *et al.* COMT val158met genotype affects μ-opioid neurotransmitter responses to a pain stressor. *Science* **299,** 1240–1243 (2003).
21. Zubieta, J. K. *et al.* Regional mu opioid receptor regulation of sensory and affective dimensions of pain. *Science* **293,** 311–315 (2001).
22. Moles, A., Kieffer, B. L. & D'Amato, F. R. Deficit in attachment behavior in mice lacking the μ-opioid receptor gene. *Science* **304,** 1983–1986 (2004).
23. Watkins, L. R. & Mayer, D. J. Organization of endogenous opiate and nonopiate pain control systems. *Science* **216,** 1185–1192 (1982).
24. Britton, K. T. & Southerland, S. Naloxone blocks 'anxiolytic' effects of neuropeptide Y. *Peptides* **22,** 607–612 (2001).
25. Li, Y., Li, J. J. & Yu, L. C. Anti-nociceptive effect of neuropeptide Y in the nucleus accumbens of rats: an involvement of opioid receptors in the effect. *Brain Res.* **940,** 69–78 (2002).
26. Caspi, A. *et al.* Role of genotype in the cycle of violence in maltreated children. *Science* **297,** 851–854 (2002).
27. Rosenberg, N. A., Li, L. M., Ward, R. & Pritchard, J. K. Informativeness of genetic markers for inference of ancestry. *Am. J. Hum. Genet.* **73,** 1402–1422 (2003).
28. Templeton, A. R. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apoprotein E locus. *Genetics* **140,** 403–409 (1995).
29. Karvonen, M. K. *et al.* Association of a leucine(7)-to-proline(7) polymorphism in the signal peptide of neuropeptide Y with high serum cholesterol and LDL cholesterol levels. *Nature Med.* **4,** 1434–1437 (1998).
30. Niskanen, L. *et al.* Leucine 7 to proline 7 polymorphism in the neuropeptide Y gene is associated with retinopathy in type 2 diabetes. *Exp. Clin. Endocrinol. Diabetes* **108,** 235–236 (2000).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** This work was supported by grants R01 DA 016423 to J.-K.Z., K01 MH072837 and a NARSAD Young Investigator Award to A.R.H., R01-AA13892, P50-DA16556 and K02-DA17232 to R.S., PO1 HL040962 and RO1 HL065137 to S.B.M., and R01 DE 15396 to C.S.S.

**Author Contributions** Z.Z. and G.Z. performed molecular biology and genetic analyses. A.R.H., R.E.F., S.B.M. and S.M.B. generated and analysed fMRI data. J.-K.Z., D.S. and C.S.S. generated and analysed PET data. R.S. and R.L.H. generated plasma NPY data. J.-K.Z., A.R.H., Z.Z., G.Z., D.S., M.V. and D.G. all contributed to conceptual issues. M.V. and D.G. were involved in the collection of the Finnish data set. D.C.M. provided post-mortem tissue samples. M.-A.E., R.H.L, X.-Z.H., C.H., K.X., B.B., Q.Y. and P.-H.S. contributed to data analysis. D.G. conceived and planned the study. Z.Z. and D.G. wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.G. (davidgoldman@mail.nih.gov).

## METHODS

**Genotyping.** *NPY* genotyping was carried out using 5′-nuclease assays for six SNP markers (rs17149106, rs16147, rs16139, rs5573, rs5574 and rs16475) in accordance with the standard TaqMan Assay protocol (ABI). A dinucleotide insertion/deletion polymorphism (rs3037354) was genotyped by sequencing because of the presence of an adjacent SNP. Results were analysed on an ABI 7900 Sequence Detector, using Sequence Detection System 2.0 software. Genotyping accuracy was verified by genotyping in a replicate 10% of samples, randomly selected. Genotyping accuracy was greater than 0.99. Genotype completion rate was greater than 0.96. Sequencing was performed with BigDye Terminator v3.1 reagent and analysed on an ABI 3100 sequencer. The Finnish Caucasian sample, the PET imaging pain stress response sample (Ann Arbor) and the sample measured for plasma NPY levels (New Haven) were also genotyped with AIMs with the use of a GoldenGate Assay (Illumina) 1,536-marker array. The marker selection was based on the HapMap Project reference allele frequency (RAF) among European Caucasians, Asians, and Africans (Yoruban). Selected markers have an RAF difference greater than 0.7 and RAF ratio larger than 10:1 between at least one pair of the continental populations and is balanced to distinguish between continental populations[31].

***NPY* mRNA quantification by real-time PCR.** Total RNA was isolated from lymphoblastoid cell lines derived from healthy Finnish men with TRIzol reagents (Invitrogen). RNA samples were treated with RNase-free DNase (Ambion) to remove contaminating DNA. *NPY* mRNA quantification was performed with a two-step TaqMan real-time RT–PCR assay. Total RNA (1 μg) was reverse-transcribed into cDNA with a Cloned AMV First-Strand cDNA Synthesis Kit and random hexamers as primers, following the manufacturer's instructions. cDNA (3 μl) was applied in 25 μl of a TaqMan real-time PCR reaction on an ABI 7700 Sequence Detector with the ABI Assay-on-demand gene expression kit and using β-actin as an endogenous control for normalization. Signals were analysed with Sequence Detector System software 2.0. Relative *NPY* mRNA expression was calculated by the formula $2^{-\Delta\Delta C_t}$, where $\Delta C_t$ is the difference in $C_t$ (PCR cycle number when signal has reached the threshold) between NPY and β-actin. A triplicate of each sample was used to determine each measurement and the assay was repeated at least three times.

**Detection of allele-specific expression of human *NPY* in cerebellum.** Differential allelic expression analysis was performed with a RT-coupled 5′ nuclease assay. A coding SNP, rs5574 (C/T) located in *NPY* exon 3, was selected as an endogenous reporter for the assay (see Supplementary Table 3 for primer/probe sequences). Total RNA (1 μg) extracted from post-mortem cerebella of individuals heterozygous at rs5574 was reverse-transcribed. cDNA (2 μl) was used in a 25-μl reaction for TaqMan real-time PCR. For the control of amplification efficiency, a template with an artificial 1:1 allelic ratio, constructed by inserting one copy each of the cDNA amplicon containing either the C allele and T allele in tandem into the pDsRed2-1 plasmid, was also assayed on the same plate. The principle and the method of computation of allele expression ratios were described previously[32].

**Quantification of plasma NPY.** Blood samples were obtained at 08:45 (45 min after placement of the intravenous catheter) and every 15 min thereafter for a total of eight time points under the neutral condition. Plasma concentrations of NPY were measured by using our previously well-characterized double-antibody radioimmunoassay with $^{125}$I-NPY as the tracer[33]. The working range of the NPY assay is 19.5–1,250 pg ml$^{-1}$, and the assay sensitivity is about 15 pg ml$^{-1}$. The intra-assay and inter-assay coefficients of variation are about 4% and 14%, respectively. NPY concentrations were standardized with the $Z$ score to minimize between-assay variation.

**Report gene constructs and transfection assays.** Human *NPY* promoter and upstream regulatory region (from −1016 bp to 63 bp) were amplified by PCR in a 50-μl volume from genomic DNA of individuals who were either homozygous or heterozygous for the five common haplotypes of the region. The PCR fragments were digested with both *Xho*I and *Hin*dIII restriction enzymes, and ligated into the multiple cloning sites upstream of the DsRed protein reporter gene of the pDsRed2-1 plasmid (BD Biosciences). Ligation products were used to transform 50 μl of TOP10 competent cells (Invitrogen). Plasmid DNAs were extracted, and clones with an insert of the correct size were verified by DNA sequencing. For transfection analysis, RN46A cells were grown in 50% DMEM medium and 50% Ham's F-12 medium supplemented with 10% FBS. Cells were plated out 24 h before transfection at a density of $5 \times 10^5$ cells per well in 24-well plates, and each group of samples was plated in triplicate wells. The constructed *NPY* promoter haplotype-pDsRed plasmid DNA (1.6 μg) and green fluorescent protein (GFP) plasmid DNA (0.05 μg), used as a transfection efficiency control, were co-transfected into the cultured cells with Lipofectamine 2000 (Invitrogen) in accordance with the manufacturer's instructions. pDsRed vector DNA without a *NPY* promoter insert was also used for each transfection as a negative control. Both red (DsRed protein reporter gene) and green (GFP) fluorescent signals were captured through a fluorescent microscope with the use of OpenLab software (Improvision), 48 h after transfection. For each well of the transfected cells, ten fields of signals were captured across the same centre areas of the well, and a total of 30 fields of both red and green signal were captured for each group of cells. For each field red and green fluorescent signals were measured in the same region of interest (ROI) with the red signal of each field normalized against the green signal. The transfection analysis was repeated at least three times. The mean and standard deviation of reporter gene signal were calculated for each promoter haplotype group, and the expression levels were compared between haplotypes by using a $t$-test (two-tailed).

**Data analysis.** Genotype distributions were tested at each polymorphic locus for departure from Hardy–Weinberg equilibrium. Pairwise linkage disequilibrium coefficients ($D'$) were estimated and linkage disequilibrium block structures were evaluated with Haploview[34]. A minimum average $D'$ value of 0.80 was used to define block boundaries (however, the average $D'$ in most blocks was substantially higher). Maximum-likelihood estimates of haplotypes and frequencies were determined with PHASE v2.02 (ref. 35), and diplotypes were assigned to each individual. For haplotype cladistic clustering, the Manhattan distances among haplotypes were calculated with a modified metric in which more weight was given to markers with higher minor allele frequency and to markers with higher $r^2$ values to neighbouring markers (Q.Y., unpublished observations). Continuous variables were compared among diplotype groups by a two-tailed $t$-test, analysis of variance, and regression analysis, and are presented as means ± s.e.m.


31. Enoch, M. A., Shen, P. H., Xu, K., Hodgkinson, C. & Goldman, D. Using ancestry-informative markers to define populations and detect population stratification. *J. Psychopharmacol.* **20,** 19–26 (2006).
32. Zhu, G. *et al.* Differential expression of human COMT alleles in brain and lymphoblasts detected by RT-coupled 5′ nuclease assay. *Psychopharmacology (Berl.)* **177,** 178–184 (2004).
33. Allen, R. *et al.* Neuropeptide Y radio-immunoassay: characterization and application. *Clin. Exp. Pharmacol. Physiol.* **18,** 825–833 (1991).
34. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).
35. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73,** 1162–1169 (2003).

Vol 452 | 24 April 2008 | doi:10.1038/nature06850

LETTERS

# Insect olfactory receptors are heteromeric ligand-gated ion channels

Koji Sato[1], Maurizio Pellegrino[2], Takao Nakagawa[1]†, Tatsuro Nakagawa[1], Leslie B. Vosshall[2] & Kazushige Touhara[1]

**In insects, each olfactory sensory neuron expresses between one and three ligand-binding members of the olfactory receptor (OR) gene family, along with the highly conserved and broadly expressed Or83b co-receptor[1–9]. The functional insect OR consists of a heteromeric complex of unknown stoichiometry but comprising at least one variable odorant-binding subunit and one constant Or83b family subunit[10–16]. Insect ORs lack homology to G-protein-coupled chemosensory receptors in vertebrates[17] and possess a distinct seven-transmembrane topology with the amino terminus located intracellularly[10,18]. Here we provide evidence that heteromeric insect ORs comprise a new class of ligand-activated non-selective cation channels. Heterologous cells expressing silkmoth, fruitfly or mosquito heteromeric OR complexes showed extracellular $Ca^{2+}$ influx and cation-non-selective ion conductance on stimulation with odorant. Odour-evoked OR currents are independent of known G-protein-coupled second messenger pathways. The fast response kinetics and OR-subunit-dependent $K^+$ ion selectivity of the insect OR complex support the hypothesis that the complex between OR and Or83b itself confers channel activity. Direct evidence for odorant-gated channels was obtained by outside-out patch-clamp recording of *Xenopus* oocyte and HEK293T cell membranes expressing insect OR complexes. The ligand-gated ion channel formed by an insect OR complex seems to be the basis for a unique strategy that insects have acquired to respond to the olfactory environment.**

To examine the mechanism of signal transduction of multimeric OR + Or83b complexes, we expressed ORs with known ligand specificities from silkmoth (*Bombyx mori*), fruitfly (*Drosophila melanogaster*) and the malaria vector mosquito (*Anopheles gambiae*), together with members of the Or83b co-receptor family from each species (BmOr-2, Or83b and GPROR7, respectively) in heterologous cells and performed $Ca^{2+}$ imaging and electrophysiology experiments. HeLa cells expressing *Drosophila* Or47a + Or83b or *Anopheles* GPROR2 + GPROR7 showed ligand-dependent and dose-dependent[7,12,19] increases in intracellular $Ca^{2+}$ that required the Or83b-class co-receptor[10–13] (Fig. 1a, b, and Supplementary Fig. 1). Inward current responses to pentyl acetate (PA) or 2-methyl phenol (2-MP) were observed by whole-cell patch-clamp recordings of HeLa cells expressing Or47a + Or83b or GPROR2 + GPROR7, respectively, at a holding potential of −60 mV (Fig. 1c, d). The current showed a slight outward rectification similar to that previously observed for *Bombyx* BmOr-1 + BmOr-2 expressed in *Xenopus* oocytes[12]. Reversal potentials of $-1.9 \pm 0.8$ and $+0.2 \pm 1.1$ mV (means ± s.e.m.) for Or47a + Or83b and GPROR2 + GPROR7, respectively, were measured (Fig. 1e–h). Waveforms of outward and inward currents were completely symmetrical (Fig. 1e, f), suggesting that odorants activate a non-selective cation conductance through the insect OR complex.

We next characterized the source of increase in intracellular $Ca^{2+}$ when insect ORs are stimulated by odour ligands (Fig. 2). Chelating

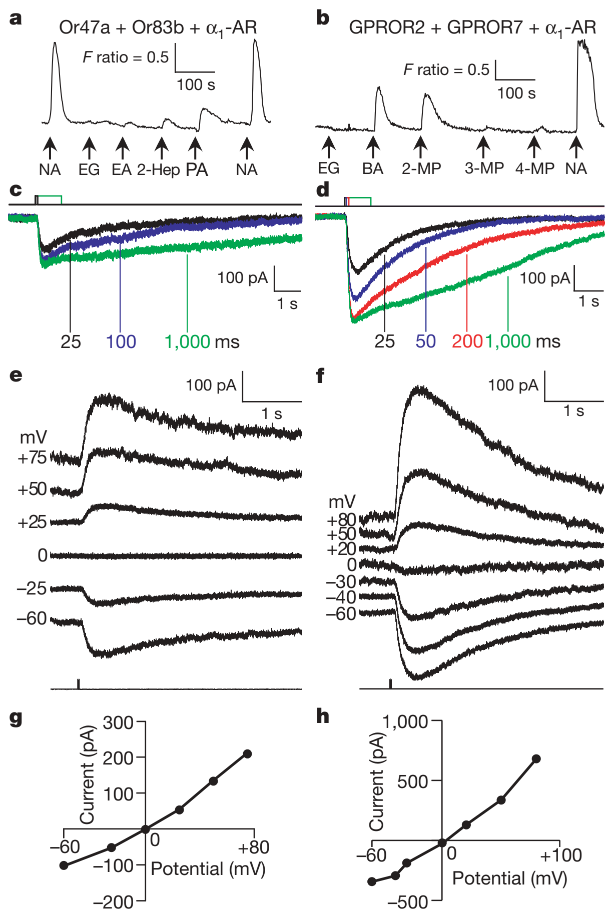



**Figure 1 | Odorants activate a non-selective cation conductance in HeLa cells expressing multimeric insect ORs. a, c, e, g,** Recording from a HeLa cell expressing Or47a, Or83b and $\alpha_1$-adrenergic receptor ($\alpha_1$-AR) on stimulation with 100 μM pentyl acetate (PA), ethyl acetate (EA), 2-heptanone (2-Hep), eugenol (EG) and 100 nM noradrenaline (NA). **b, d, f, h,** Recording from a HeLa cell expressing GPROR2, GPROR7 and $\alpha_1$-AR on stimulation with 100 μM 2-methyl phenol (2-MP), 3-methyl phenol (3-MP), 4-methyl phenol (4-MP), EG, benzaldehyde (BA) and 100 nM NA. **a, b,** Fura-2 acetoxymethyl ester (Fura-2/AM) $Ca^{2+}$ imaging in response to ligand stimulation for 10 s (arrows indicate onset). **c, d,** Inward current responses with various stimulus durations of PA (**c**) and 2-MP (**d**). **e, f,** Current responses to stimulus (bottom traces indicate onset) at various holding potentials ($n = 14$ and 4, respectively). **g, h,** Current–voltage (*I–V*) relationship; the respective peak current of each response in **e** and **f** is plotted.

[1]Department of Integrated Biosciences, The University of Tokyo, Chiba 277-8562, Japan. [2]Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, New York 10065, USA. †Present address: Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, New York 10065, USA.

extracellular $Ca^{2+}$ with EGTA or bis-(*o*-aminophenoxy)ethane-*N*,*N*,*N*′,*N*′-tetra-acetic acid (BAPTA) strongly diminished the increase in $Ca^{2+}$ mediated by Or47a + Or83b odour activation but had no effect on the increase in $Ca^{2+}$ stimulated by activation of the $G\alpha_q$-coupled $\alpha_1$-adrenergic receptor ($\alpha_1$-AR) by noradrenaline (Fig. 2a, b). This suggests that most of the increase in intracellular $Ca^{2+}$ observed when insect ORs are activated derives from the influx of extracellular $Ca^{2+}$. *Drosophila* olfactory sensory neurons show receptor-dependent spontaneous activity in the absence of odour ligands[4,7,8,20,21], and we examined whether such spontaneous activity can be detected in heterologous cells expressing insect ORs. In Or47a + Or83b-expressing cells, the resting intracellular $Ca^{2+}$ level was higher than that in mock-transfected cells (Supplementary Fig. 2a) and this was strongly dependent on the extracellular $Ca^{2+}$ concentration (Supplementary Fig. 2b, c). Basal intracellular $Ca^{2+}$ levels were decreased by extracellular application of EGTA in Or47a + Or83b-expressing cells (Fig. 2a and Supplementary Fig. 2b), suggesting that insect ORs mediate spontaneous $Ca^{2+}$ influx in the absence of odour ligand.

To examine whether this odour-evoked activation requires known intracellular signalling cascades, we performed pharmacological experiments. No current response was observed in whole-cell patch-clamp experiments using a recording pipette filled with intracellular solution containing cyclic AMP, cyclic GMP or inositol-1,4,5-trisphosphate ($IP_3$) (Supplementary Fig. 3a). U73122, an inhibitor of phospholipase C, abolished $\alpha_1$-AR-mediated $Ca^{2+}$ responses but did not affect increases in $Ca^{2+}$ mediated by Or47a + Or83b (Fig. 2b, d). OR complex activation is therefore largely independent of $G\alpha_q$/PLC signalling (Fig. 2b, d). Moreover, insect OR activation fails to stimulate the $G\alpha_s$/cAMP pathway, because we did not detect increases in intracellular cAMP in HEK293T cells expressing Or47a + Or83b, BmOr-1 + BmOr-2 or GPROR2 + GPROR7 after stimulation with odour (Fig. 2c and Supplementary Fig. 4a). Control cells expressing the mouse OR, mOR-EG, showed robust increases. cAMP produced by stimulation of mOR-EG activated cyclic nucleotide-gated channel A2 (CNGA2), but activation of Or47a + Or83b with pentyl acetate had no effect on this channel (Supplementary Fig. 4b, c). Application of 8-bromo-cGMP or 8-bromo-cAMP failed to produce $Ca^{2+}$ responses in HEK293T cells or oocytes expressing Or47a + Or83b or GPROR2 + GPROR7 (Supplementary Fig. 5). Although we failed to detect activation of a $G\alpha_s$/cAMP pathway by ligand stimulation of Or47a + Or83b, BmOr-1 + BmOr-2 or GPROR2 + GPROR7, we did observe that some OR complexes showed a small, ligand-independent sensitivity to cyclic nucleotides (Supplementary Fig. 5b). The basis of this sensitivity was not investigated further in this study. To examine whether G-protein signalling was necessary for insect OR activation, we performed experiments with GDP-βS, a non-hydrolysable form of GDP that inhibits G-protein-coupled signalling. GDP-βS had no effect on the current response of BmOr-1 + BmOr-2 to bombykol (Fig. 2e) or the response of Or47a + Or83b to pentyl acetate (Supplementary Fig. 3d) when these receptor complexes were expressed in oocytes, but GDP-βS significantly inhibited $Ca^{2+}$-activated $Cl^-$ current responses induced by the activation by noradrenaline of $\alpha_1$-AR/$G\alpha_q$/PLC- and cystic fibrosis transmembrane conductance regulator (CFTR) currents elicited by stimulation of $\beta_2$-adrenergic receptor ($\beta_2$-AR)/$G\alpha_s$/adenylyl cyclase pathways by isoprenaline (isoproterenol) (Fig. 2e). These results make it unlikely that G-protein-signalling cascades underlie responses of the insect OR complex.

To examine the response kinetics of an OR heteromultimer, we recorded increases in intracellular $Ca^{2+}$ and electrical responses in the same cell. A HeLa cell expressing Or47a + Or83b and $\alpha_1$-AR showed a $Ca^{2+}$ response to both pentyl acetate and noradrenaline, but showed an electrical response to pentyl acetate only, not to noradrenaline (Fig. 3a). The latency of the Or47a + Or83b-mediated $Ca^{2+}$ response was less than one-tenth that of the $\alpha_1$-AR/$G\alpha_q$/PLC-mediated $Ca^{2+}$ response ($240 \pm 46$ ms versus $3{,}300 \pm 400$ ms). The latencies of current responses of Or47a + Or83b (Fig. 3b) and GPROR2 + GPROR7 (Fig. 3c) were $17.9 \pm 3.1$ and $28.5 \pm 1.9$ ms, respectively. This is considerably shorter than the about 50–200-ms latency of the G-protein-coupled olfactory response in vertebrate olfactory neurons[22,23], in which the loading of G proteins is the rate-limiting step[23]. The time course of $Ca^{2+}$ responses of Or47a + Or83b and GPROR2 + GPROR7 was also distinct from those of $\alpha_1$-AR (Fig. 3f) and mOR-EG (Fig. 3g). The slope of the odorant-induced insect OR response was not modulated by the duration of stimulation (Fig. 3d, e), whereas the slope of noradrenaline-induced $\alpha_1$-AR or eugenol-induced mOR-EG responses was sharper when the stimulation period was prolonged (Fig. 3f, g).

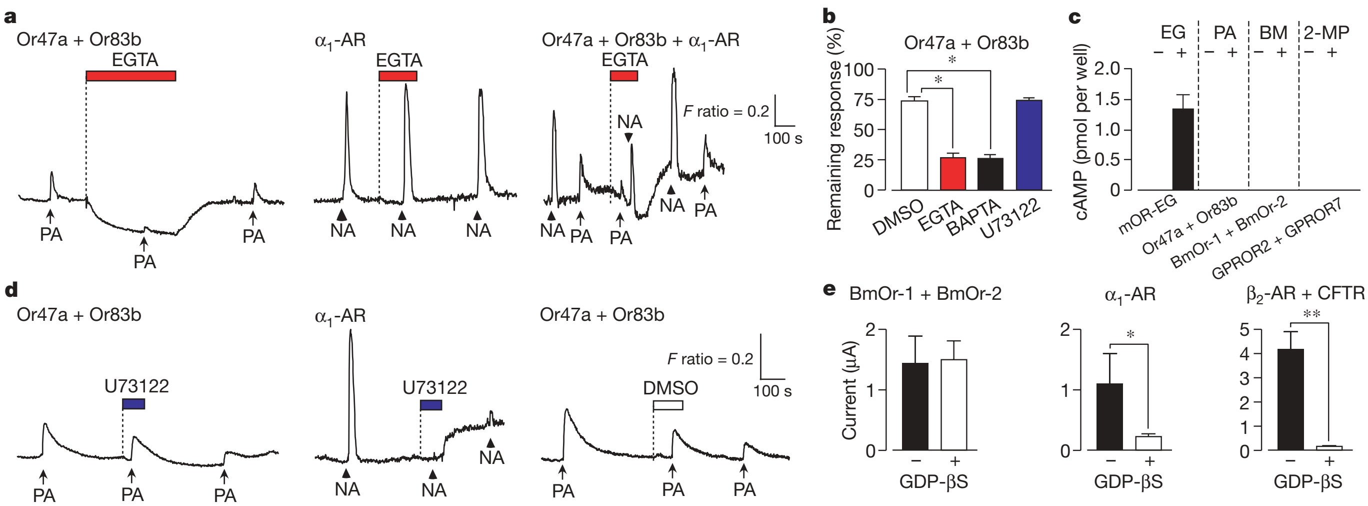



**Figure 2 | Insect OR activity is independent of G protein signalling. a**, $Ca^{2+}$ responses of HeLa cells expressing Or47a + Or83b, $\alpha_1$-AR or Or47a + Or83b + $\alpha_1$-AR to a 10-s stimulation with 100 μM pentyl acetate (PA) or 100 nM noradrenaline (NA) with application of 10 mM EGTA (red bar). **b**, Effects of $Ca^{2+}$ chelators (10 mM EGTA and 5 mM BAPTA) and U73122 (PLC inhibitor; containing 0.1% DMSO) on Or47a + Or83b $Ca^{2+}$ response (DMSO control, $n = 28$; U73122, $n = 18$; EGTA, $n = 10$; BAPTA, $n = 20$). Significance assessed by *t*-test: asterisk, $P < 0.001$. **c**, cAMP production in HEK293T cells expressing mOR-EG, Or47a + Or83b, BmOr-1 + BmOr-2 or GPROR2 + GPROR7 stimulated with eugenol (EG; 1 mM), PA (100 μM), bombykol (BM; 10 μM) or 2-methyl phenol (2-MP; 100 μM), respectively. **d**, As in **a**, but with application of 10 μM U73122 (filled bar) or 0.1% DMSO (open bar). **e**, Effect of GDP-βS on ligand-induced inward currents in *Xenopus* oocytes expressing BmOr-1 + BmOr-2 (30 μM BM), $\alpha_1$-AR (1 μM NA) or $\beta_2$-AR + CFTR (10 μM isoprenaline). Significance assessed by *t*-test: asterisk, $P < 0.05$; two asterisks, $P < 0.01$; $n = 5$ each. Data are shown as means ± s.e.m.

If the heteromultimer combination of the insect OR complex determines ion selectivity, we might expect that the ion permeability of odour-gated currents would differ with different subunit compositions. Indeed, both the reversal potential and outward rectification of Or47a + Or83b and Or47a + BmOr-2 differed slightly (Supplementary Fig. 6a, b). Ruthenium red, a cation-channel blocker, which completely inhibits odour-evoked inward current responses of BmOr-1 + BmOr-2 in *Xenopus* oocytes[12], inhibited both bombykol-evoked inward current and basal $Ca^{2+}$ levels (Supplementary Fig. 6c, d) of BmOr-1 + BmOr-2-expressing cells but had no effect on HEK293T cells expressing GPROR2 + GPROR7 (data not shown). This suggests that insect OR subunit composition affects the degree of inhibition by ruthenium red.

To obtain evidence that currents are produced directly by the insect ORs, outside-out patch-clamp recordings were performed on oocyte membranes containing Or47a + Or83b or GPROR2 + GPROR7. First we established that oocytes expressing GPROR2 + GPROR7 showed high-affinity ligand-specific macroscopic currents that depended on the expression of both subunits (Supplementary Fig. 7a) and required extracellular $Na^{+}$ (Supplementary Fig. 7b, d). Patches from uninjected control oocytes did not show current responses to two different odour stimuli or oocyte Ringer's solution (Supplementary Fig. 7e), indicating that odorants do not induce non-specific currents by interacting with oocyte membrane or membrane proteins. Outside-out patch preparations of oocytes injected with Or47a + Or83b or GPROR2 + GPROR7 showed transient unitary currents whose frequency increased with the application of the cognate ligands pentyl acetate or 2-methyl phenol (Fig. 4 and Supplementary Fig. 7f–h). The channel conductance of the Or47a + Or83b complex was 1.2 ± 0.03 pA ($n = 4$) at a holding potential of −80 mV (Fig. 4d). The magnitude of odour-evoked currents was voltage-dependent and increased when the holding potential was progressively changed from +40 mV to −120 mV (Supplementary Fig. 7h). The frequency of these odour-evoked events was concentration-dependent (Fig. 4e). There was no difference in pentyl-acetate-evoked conductance from membranes expressing Or47a + Or83b in the presence or absence of ATP (1 mM)/GTP (100 μM) in the recording pipette, suggesting that ATP/GTP-mediated signal transduction components are not involved in the primary conductance of this channel (without ATP/GTP shown in Fig. 4a–c; with ATP/GTP shown in Supplementary Fig. 3b, c).

Similar results were obtained in outside-out patch-clamp recording of HEK293T membranes expressing Or47a + Or83b or GPROR2 + GPROR7 (Fig. 5). In contrast with the bath application

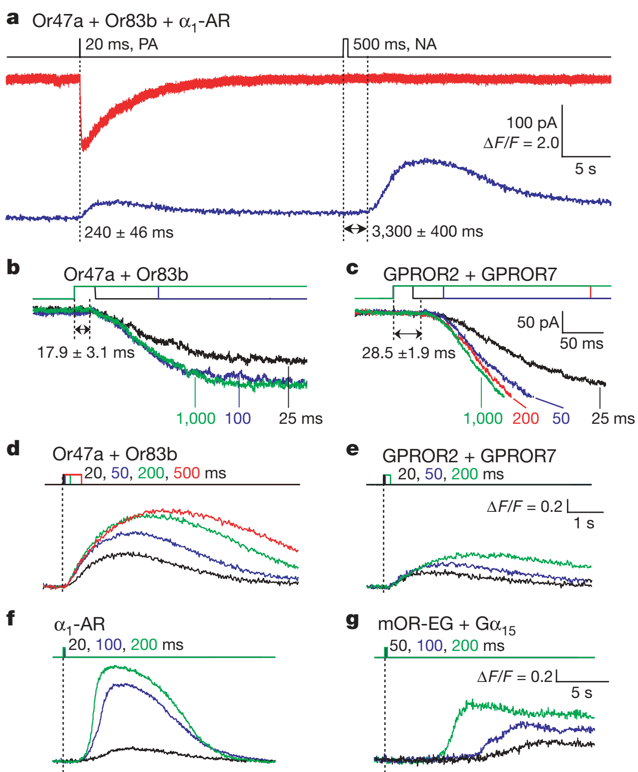



**Figure 3 | Kinetic analysis of $Ca^{2+}$ and current responses of insect ORs in HeLa cells.** **a**, Simultaneous recording of whole-cell current (red trace) and intracellular $Ca^{2+}$ (Fluo-4-4K; blue trace) from a HeLa cell expressing Or47a + Or83b + $\alpha_1$-AR. Stimulus durations of pentyl acetate (PA; 20 ms) and noradrenaline (NA; 500 ms) are indicated at the top. $Ca^{2+}$ response latency is shown as mean ± s.e.m. ($n = 8$). **b**, **c**, Onset of inward current responses of Or47a + Or83b to PA (**b**) and GPROR2 + GPROR7 to 2-methyl phenol (2-MP) (**c**) applied with various stimulus durations as indicated by the different coloured curves. The latency of current response is shown as mean ± s.e.m.; scales for **b** and **c** are shown at the right in **c** (**b**, $n = 14$; **c**, $n = 4$). **d**–**g**, Onset of Fluo-4-acetoxymethyl ester (Fluo-4/AM)-based $Ca^{2+}$ responses of Or47a + Or83b to PA (**d**), GPROR2 + GPROR7 to 2-MP (**e**), $\alpha_1$-AR to NA (**f**), and mOR-EG along with $G\alpha_{15}$ to EG (**g**), with stimuli applied for various durations as indicated. Scales for **d** and **e** are shown at the right in **e**, and for **f** and **g** at the right in **g**.

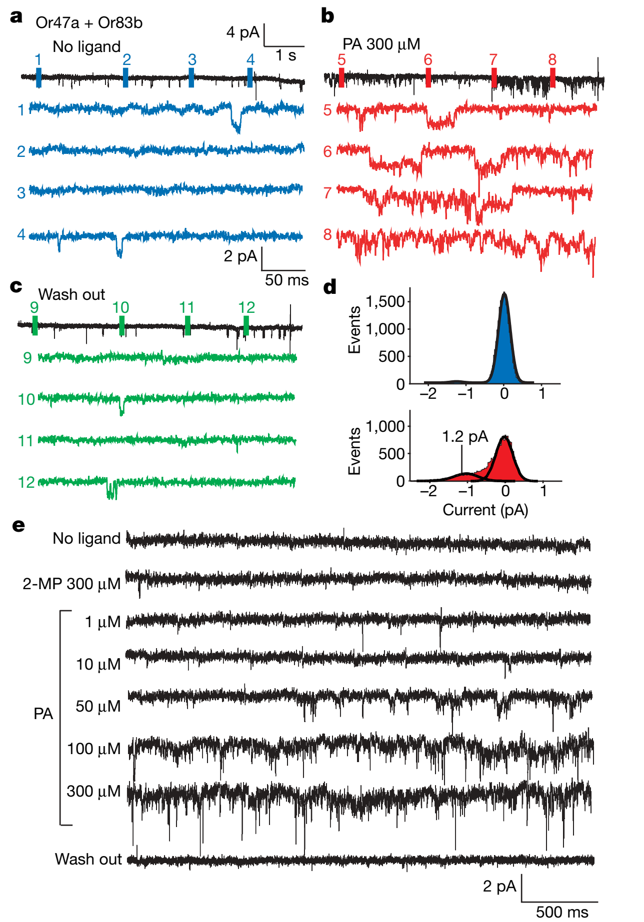



**Figure 4 | Excised outside-out patch-clamp recording of Or47a + Or83b currents measured in oocyte membranes.** **a**–**c**, Outside-out patch-clamp recording of a *Xenopus* oocyte membrane expressing Or47a + Or83b before stimulation (**a**), during stimulation with 300 μM pentyl acetate (PA) (**b**), and after wash out (**c**). The bottom traces of each panel indicate expansions of 300-ms current traces of single-channel recording at the positions indicated by the numbers. The data for **a**–**c** were obtained from the same cell with voltage clamped at −80 mV. Scales for **a**–**c** are indicated at the top in **a** and scales for the expansions are at the bottom in **a**. **d**, All-point current histograms of unitary events before (blue) and during (red) application of the ligand PA in **b** and **c**. Amplitude distributions were fitted with two gaussian components (black lines). **e**, Dose–response and ligand selectivity of Or47a + Or83b currents with voltage clamped at −90 mV.

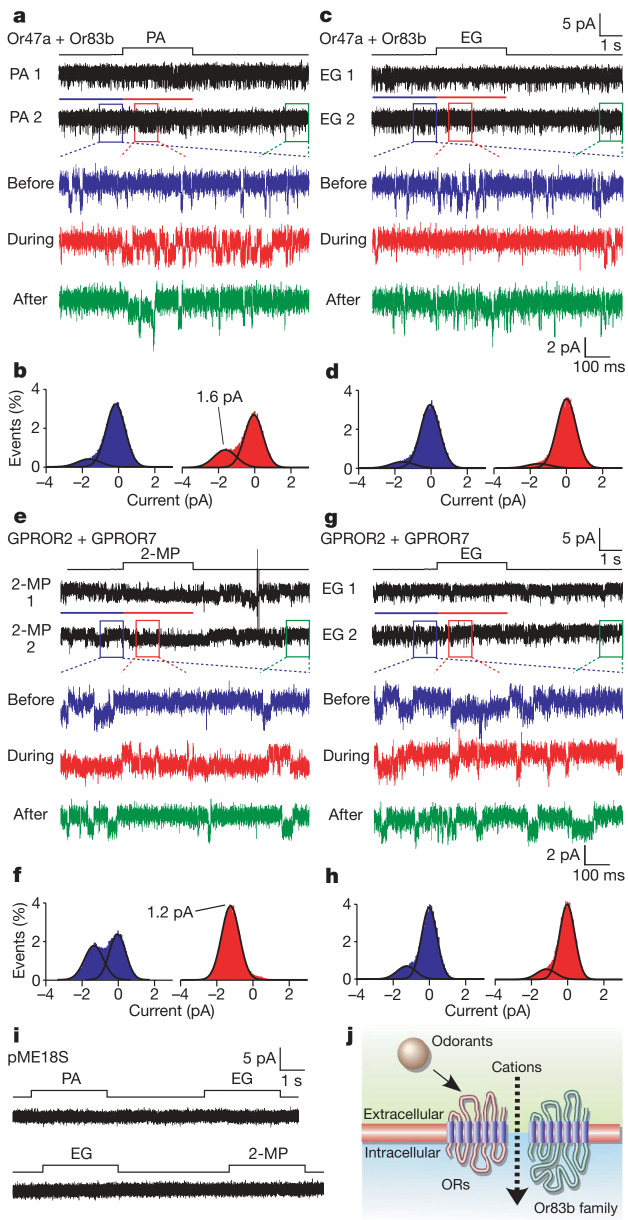



**Figure 5 | Excised outside-out patch-clamp recording of insect OR currents measured in HEK293T membranes. a–d**, Or47a + Or83b-expressing HEK293T membranes stimulated for 3 s with pentyl acetate (PA; 300 μM) (**a**, **b**) or the non-agonist eugenol (EG; 300 μM) (**c**, **d**). The bottom traces show expansions of the current trace of the single-channel recording at the positions indicated by coloured boxes on trace 2 before (blue), during (red) and after (green) stimulation with PA (**a**) or EG (**c**). All-point current histograms were obtained from the region indicated on trace 1 by the blue bar without stimulation (blue histogram, left) and by the red bar for stimulation with PA (**b**) or EG (**d**) (red histogram, right). Mean peak current level (1.6 pA) was obtained from the fitted gaussians. Data in **a–d** were obtained from the same cell. Scales for **a** and **c** are indicated at the top in **c** and scales for the expansions are at the bottom in **c**. The scale on the *y* axis indicates the relative number of current events with a bin width of 0.05 pA. **e–h**, As in **a–d**, but using GPROR2 + GPROR7-expressing HEK293T membranes stimulated for 3 s with 2-methyl phenol (2-MP; 300 μM) (**e**, **f**) or the non-agonist EG (300 μM) (**g**, **h**) Scales for **e** and **g** are indicated at the top in **g** and scales for the expansions are at the bottom in **g**. Data in **e–h** were obtained from the same cell. **i**, Outside-out patch-clamp recording of vector-transfected HEK293T membranes stimulated for 3 s with 300 μM PA, EG and 2-MP. Membranes in **a–i** were at a holding potential of −60 mV. **j**, Schematic model for ligand-gated channel properties of the OR + Or83b complex.

used for oocyte experiments in Fig. 4, we used a puff application of odour ligands. Pentyl acetate evoked unitary currents of 1.5 ± 0.06 pA ($n = 4$) in outside-out patch membranes expressing Or47a + Or83b and held at −60 mV (Fig. 5a, b, and Supplementary Fig. 8a–c). We calculated a slope conductance of 27 pS for Or47a + Or83b from a current–voltage (*I–V*) curve (Supplementary Fig. 8b). Stimulation of Or47a + Or83b-expressing cell membrane with the non-ligand eugenol did not increase currents (Fig. 5c, d). The same experiments performed in membranes expressing GPROR2 + GPROR7 produced unitary currents of 1.2 ± 0.07 pA ($n = 3$) with a conductance of 20 pS at −60 mV in response to the cognate ligand 2-methyl phenol (Fig. 5e, f, and Supplementary Fig. 8d, e). As seen for Or47a + Or83b, the non-ligand eugenol did not stimulate GPROR2 + GPROR7 currents (Fig. 5g, h). The presence of currents in the absence of cognate ligand suggests that both Or47a + Or83b and GPROR2 + GPROR7 are spontaneously active, which is consistent with our other results in Figs 2a, 4a and Supplementary Figs 3 and 6d. Untransfected HEK293T membranes did not show currents on stimulation with pentyl acetate, eugenol or 2-methyl phenol (Fig. 5i). In both oocytes (Fig. 4) and mammalian cells (Fig. 5), single-channel events predominated. We attribute this either to low expression levels of these insect ORs in heterologous cell membrane patches or to unstable characteristics of the multimeric receptor in these cells. Nevertheless, these outside-out patch-clamp experiments using two different membrane preparations provide initial evidence that the insect OR complex possesses ligand-gated channel properties.

Taken together, these data provide compelling evidence that a heteromeric complex of a conventional insect OR and the highly conserved Or83b family co-receptor has the characteristics of a cation non-selective ion channel directly gated by odour or pheromone ligands (Fig. 5j). We conclude that G-protein-mediated signalling is negligible in producing the current elicited by odour activation of insect OR heteromultimers. Our findings provide insight into long-argued insect olfactory transduction mechanisms and may explain the lack of clear consensus on the role of second messengers in this process[24]. The insect ORs share no homology with any previously described ion channel and do not contain any known ion selectivity filter motifs[25]. We observed that insect OR activity was not inhibited by $Gd^{3+}$ (data not shown), a lanthanide that is a broad-spectrum ion channel inhibitor[26–28]. Therefore, although the ionic permeability reported here for $Na^+$, $K^+$ and $Ca^{2+}$ would be consistent with the properties of non-selective cation channels, a molecular basis for this novel ionotropic activity remains to be elucidated. The spontaneous activity of the OR complex found here seems to account for previous observations that olfactory sensory neurons exhibit bipolar electrical activity and become electrically negative on the deletion of Or83b *in vivo*[4,7,8]. Given that there are 62 and 79 potential ligand-binding OR subunits in *Drosophila* and *Anopheles*, respectively, the insect ORs may represent the largest single family of ion-channel-like proteins in any organism. Our work also raises the intriguing possibility that the insect gustatory system, which senses bitter and sweet tastants as well as carbon dioxide, shares this ionotropic coupling mechanism with the insect ORs. In fact, an ionotropic sugar-gated channel in fleshfly taste cells has previously been reported[29]. Our finding offers the caveat that other orphan receptors classified as G-protein-coupled receptors merely because of their putative seven-transmembrane topology may instead possess ligand-gated channel activities, as has been shown previously for light-activated channelrhodopsin[30]. This work has important implications for worldwide efforts to identify specific inhibitors for the insect ORs, which may prove useful in controlling host-seeking behaviours of disease-vector insects such as mosquitoes.

*Note added in proof:* The authors of an accompanying paper[31] also conclude that insect ORs possess ion channel-like properties, but they report $G_s$-dependent cyclic-nucleotide activation of Or83b alone, which we fail to observe.

## METHODS SUMMARY

**Odorant stocks and stimulation.** Stock solutions of bombykol, isoprenaline, forskolin, 3-isobutyl-1-methylxanthine, ruthenium red, and U73122 were prepared in dimethylsulphoxide (DMSO). Other stimuli in Figs 1–3 and 5 were added directly to Ringer's solution. Stimuli in Fig. 4 were prepared in DMSO as 1 M stocks, which were then diluted into the extracellular solution. For patch-clamp recording and kinetic analysis, odorants were applied focally to the recording cell with the use of a handmade pressure-ejection system and a TTL (transistor–transistor logic) pulse generator. For other experiments, odorants and inhibitors were applied by perfusing the bath solution.

**Electrophysiology and $Ca^{2+}$ imaging.** Two-electrode voltage-clamp recording from *Xenopus laevis* oocytes was performed as described previously[12]. $Ca^{2+}$ imaging from mammalian cell lines was performed as described previously[32].

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 22 July 2007; accepted 20 February 2008.**
**Published online 13 April 2008.**

1. Clyne, P. J. *et al.* A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila. Neuron* **22,** 327–338 (1999).
2. Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96,** 725–736 (1999).
3. Couto, A., Alenius, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr. Biol.* **15,** 1535–1547 (2005).
4. Dobritsa, A. A. *et al.* Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37,** 827–841 (2003).
5. Elmore, T., Ignell, R., Carlson, J. R. & Smith, D. P. Targeted mutation of a *Drosophila* odor receptor defines receptor requirement in a novel class of sensillum. *J. Neurosci.* **23,** 9906–9912 (2003).
6. Fishilevich, E. & Vosshall, L. B. Genetic and functional subdivision of the *Drosophila* antennal lobe. *Curr. Biol.* **15,** 1548–1553 (2005).
7. Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125,** 143–160 (2006).
8. Hallem, E. A., Ho, M. G. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117,** 965–979 (2004).
9. Goldman, A. L., Van der Goes van Naters, W., Lessing, D., Warr, C. G. & Carlson, J. R. Coexpression of two functional odor receptors in one neuron. *Neuron* **45,** 661–666 (2005).
10. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo. PLoS Biol.* **4,** e20 (2006).
11. Larsson, M. C. *et al. Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43,** 703–714 (2004).
12. Nakagawa, T., Sakurai, T., Nishioka, T. & Touhara, K. Insect sex-pheromone signals mediated by specific combinations of olfactory receptors. *Science* **307,** 1638–1642 (2005).
13. Neuhaus, E. M. *et al.* Odorant receptor heterodimerization in the olfactory system of *Drosophila melanogaster. Nature Neurosci.* **8,** 15–17 (2005).
14. Jones, W. D., Nguyen, T. A., Kloss, B., Lee, K. J. & Vosshall, L. B. Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr. Biol.* **15,** R119–R121 (2005).
15. Krieger, J., Klink, O., Mohl, C., Raming, K. & Breer, H. A candidate olfactory receptor subtype highly conserved across different insect orders. *J. Comp. Physiol. [A]* **189,** 519–526 (2003).
16. Pitts, R. J., Fox, A. N. & Zwiebel, L. J. A highly conserved candidate chemoreceptor expressed in both olfactory and gustatory tissues in the malaria vector *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA* **101,** 5058–5063 (2004).
17. Wistrand, M., Kall, L. & Sonnhammer, E. L. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Sci.* **15,** 509–521 (2006).
18. Lundin, C. *et al.* Membrane topology of the *Drosophila* OR83b odorant receptor. *FEBS Lett.* **581,** 5601–5604 (2007).
19. Hallem, E. A., Nicole Fox, A., Zwiebel, L. J. & Carlson, J. R. Olfaction: mosquito receptor for human-sweat odorant. *Nature* **427,** 212–213 (2004).
20. de Bruyne, M., Clyne, P. J. & Carlson, J. R. Odor coding in a model olfactory organ: the *Drosophila* maxillary palp. *J. Neurosci.* **19,** 4520–4532 (1999).
21. de Bruyne, M., Foster, K. & Carlson, J. R. Odor coding in the *Drosophila* antenna. *Neuron* **30,** 537–552 (2001).
22. Sato, K. & Suzuki, N. The contribution of $Ca^{2+}$-activated $Cl^{-}$ conductance to amino-acid-induced inward current responses of ciliated olfactory neurons of the rainbow trout. *J. Exp. Biol.* **203,** 253–262 (2000).
23. Firestein, S., Shepherd, G. M. & Werblin, F. S. Time course of the membrane current underlying sensory transduction in salamander olfactory receptor neurones. *J. Physiol. (Lond.)* **430,** 135–158 (1990).
24. Kaissling, K. E. Peripheral mechanisms of pheromone reception in moths. *Chem. Senses* **21,** 257–268 (1996).
25. Zagotta, W. N. Membrane biology: permutations of permeability. *Nature* **440,** 427–429 (2006).
26. Lacampagne, A., Gannier, F., Argibay, J., Garnier, D. & Le Guennec, J. Y. The stretch-activated ion channel blocker gadolinium also blocks L-type calcium channels in isolated ventricular myocytes of the guinea-pig. *Biochim. Biophys. Acta* **1191,** 205–208 (1994).
27. Kanzaki, M. *et al.* Molecular identification of a eukaryotic, stretch-activated nonselective cation channel. *Science* **285,** 882–886 (1999).
28. Halaszovich, C. R., Zitt, C., Jungling, E. & Luckhoff, A. Inhibition of TRP3 channels by lanthanides. Block from the cytosolic side of the plasma membrane. *J. Biol. Chem.* **275,** 37423–37428 (2000).
29. Murakami, M. & Kijima, H. Transduction ion channels directly gated by sugars on the insect taste cell. *J. Gen. Physiol.* **115,** 455–466 (2000).
30. Nagel, G. *et al.* Channelrhodopsin-1: a light-gated proton channel in green algae. *Science* **296,** 2395–2398 (2002).
31. Wicher, D. *et al. Drosophila* odorant receptors are both ligand-gated and cyclic nucleotide-activated cation channels. *Nature* doi:10.1038/nature06861 (this issue).
32. Katada, S., Nakagawa, T., Kataoka, H. & Touhara, K. Odorant response assays for a heterologously expressed olfactory receptor. *Biochem. Biophys. Res. Commun.* **305,** 964–969 (2003).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** K.S. and K.T. thank M. Tominaga, Y. Okamura and Y. Kubo for discussion. M.P. and L.B.V. thank D. Gadsby, and Gadsby laboratory members N. Vedovato, P. Artigas, P. Hoff and A. Kovacs. DNA clones were provided by R. Y. Tsien (mRFP), T.-Y. Chen (CNGs) and A. Kovacs (CFTR). This work was supported in part by grants from PROBRAIN and MEXT of Japan to K.T., by a grant from the Foundation for the National Institutes of Health to R. Axel and L.B.V. through the Grand Challenges in Global Health Initiative, and by a National Institutes of Health (NIH) grant to L.B.V., a postdoctoral fellowship from the Japan Society for the Promotion of Science (JSPS) to T.N., and an NIH US–Japan Brain Research Cooperative Program grant and Japan-US Cooperative Science Program funding from the JSPS to K.T. and L.B.V.

**Author Contributions** Experiments were performed by K.S. (Figs 1, 2a–d, 3 and 5, and Supplementary Figs 1, 2, 3a, 4, 5a, b, 6 and 8), M.P. (Fig. 4 and Supplementary Figs 3b–d, 5c and 7), Takao N. (Fig. 2e) and Tatsuro N. (Supplementary Fig. 5b). K.T. and L.B.V. conceived and supervised the experiments, and wrote the paper with K.S. and M.P.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.T. (touhara@k.u-tokyo.ac.jp).

## METHODS

**Patch-clamp experiments in mammalian cell lines.** Full-length complementary DNAs for silkmoth ORs (BmOr-1 and BmOr-2), mosquito ORs (GPROR1, GPROR2 and GPROR7) and fruitfly ORs (Or47a and Or83b) were cloned into the pME18S vector. ORs were transiently transfected into HeLa or HEK293T cells with Lipofectamine 2000 reagent (Invitrogen). $\alpha_1$-AR and GFP or mRFP were co-transfected as a control. Whole-cell currents were amplified with a patch-clamp amplifier (CEZ-2400; Nihon-koden), and were digitized with PowerLab (AD Instruments). The extracellular solutions (normal Ringer's solution) contained (in mM): 140 NaCl, 5.6 KCl, 5 HEPES, 2.0 pyruvic acid sodium salt, 1.25 $KH_2PO_4$, 2.0 $CaCl_2$, 2.0 $MgCl_2$, 9.4 D-glucose (pH 7.4). The electrode solution contained (in mM): 140 KCl, 10 HEPES, 5 EGTA-2K, 10 D-glucose (pH 7.4). To record the shift in reversal potential and in equilibrium potential for a specific ion, the following external (bath) and internal (electrode) solution was used: low-$Na^+$ (*N*-methyl-D-glutamine; NMDG) external solution (in mM: 190 NMDG, 40 HEPES, 5.6 KCl, 2.0 pyruvic acid sodium salt, 1.25 $KH_2PO_4$, 2.0 $CaCl_2$, 2.0 $MgCl_2$, 9.4 D-glucose; pH 7.4); low-$K^+$ (NMDG) internal solution (in mM: 202 NMDG, 40 HEPES, 5 EGTA-2K, 10 D-glucose; pH 7.4). For the outside-out recording, normal Ringer's solution was used for the electrode solution. The data were sampled at 20 kHz and filtered at 2 kHz.

***Xenopus* oocyte electrophysiology.** Oocytes were microinjected with 25 ng of complementary RNA (cRNA) for a conventional OR and 25 ng of cRNA for the Or83b family. Whole-cell currents were recorded with a two-electrode voltage-clamp filled with 3 M KCl, and were amplified with an OC-725C amplifier (Warner Instruments), low-pass filtered at 50 Hz and digitized at 1 kHz. Outside-out patch-clamp recordings were performed 18–26 h after injection. After removal of the vitelline layer, oocytes were transferred to a Petri dish with a bath solution of oocytes Ringer's solution containing (in mM): 82.5 NaCl, 2 KCl, 1 $MgCl_2$, 5 HEPES, 1.8 $CaCl_2$ (pH 7.5). Pipettes (4–7 M$\Omega$) were covered with Sylgard (Dow Chemical Company) and filled with intracellular solution containing (in mM): 100 KOH, 10 HEPES, 1 EGTA, 100 sulphamic acid (pH 7.6). After seal formation (more than 5 G$\Omega$), patches were excised and transferred to the recording chamber, where the extracellular side was continuously superfused with extracellular solution containing (in mM): 100 NaOH, 10 HEPES, 1 $MgCl_2$, 100 sulphamic acid (pH 7.5); this was supplemented, where indicated, by the odorants. Solutions were switched by computer-driven electric valves (General Valve Corp.). Currents were recorded with an Axopatch 200A amplifier (Axon Instruments, Inc.), low-passed at 1 kHz (eight-pole Bessel; Frequency Devices), digitized at 10 kHz by means of an ITC-16 interface (Instrutech Corporation) and saved to a PC hard disk with PULSE v8.11 acquisition software (HEKA Elektronic). Data were analysed with Clampfit 9.0 (Axon Instruments, Inc.) and Origin PRO 7 (Origin Lab). The *I–V* curves showing ion permeability were produced with low-$Na^+$ solution ($Na^+$ in oocyte Ringer's solution replaced by the impermeable cation $NMDG^+$) and $Cl^-$-free solution ($Cl^-$ in oocyte Ringer's solution replaced by sulphamic acid). These experiments used bath application of ligands, precluding any measurement of the response latency of these currents.

**$Ca^{2+}$ imaging.** Insect ORs were transfected into HeLa or HEK293T cells, which were loaded with 2.5 μM Fura-2/AM or Fluo-4/AM for 30 min. Fluorescence was measured with an Aquacosmos $Ca^{2+}$ imaging system (Hamamatsu Photonics). For the simultaneous recording of whole-cell currents and intracellular $Ca^{2+}$ levels, the electrode solution (in mM: 140 KCl, 10 HEPES, 5 EGTA-2K, 10 D-glucose, 0.05 Fluo-4-pentapotassium salt (Fluo-4-4K); pH 7.4) was used. To synchronize the patch-clamp amplifier, odorant stimulator and charge-coupled device camera (ORCA-ER; Hamamatsu), all devices were under the control of a peripheral interface controller (PIC16F877; Microchip Technology, Inc.).

**cAMP assay.** HEK293T cells transfected with mOR-EG, Or47a + Or83b, BmOr-1 + BmOr-2 or GPROR2 + GPROR7 were incubated with 1 mM 3-isobutyl-1-methylxanthine for 30 min. The cells were then exposed to the indicated concentration of odorant solution for 15 min. cAMP levels were determined with an enzyme-linked immunosorbent assay kit (Applied Biosystems) in accordance with the manufacturer's directions.

Vol 452 | 24 April 2008 | doi:10.1038/nature06861



LETTERS

# *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels

Dieter Wicher[1], Ronny Schäfer[1], René Bauernfeind[1], Marcus C. Stensmyr[1], Regine Heller[2], Stefan H. Heinemann[3] & Bill S. Hansson[1]

**From worm to man, many odorant signals are perceived by the binding of volatile ligands to odorant receptors[1] that belong to the G-protein-coupled receptor (GPCR) family[2]. They couple to heterotrimeric G-proteins, most of which induce cAMP production[3]. This second messenger then activates cyclic-nucleotide-gated ion channels to depolarize the olfactory receptor neuron, thus providing a signal for further neuronal processing. Recent findings, however, have challenged this concept of odorant signal transduction in insects, because their odorant receptors, which lack any sequence similarity to other GPCRs[4], are composed of conventional odorant receptors (for example, Or22a), dimerized with a ubiquitously expressed chaperone protein[5], such as Or83b in *Drosophila*[6]. Or83b has a structure akin to GPCRs, but has an inverted orientation in the plasma membrane[4,7]. However, G proteins are expressed in insect olfactory receptor neurons[8], and olfactory perception is modified by mutations affecting the cAMP transduction pathway[9]. Here we show that application of odorants to mammalian cells co-expressing Or22a and Or83b results in non-selective cation currents activated by means of an ionotropic and a metabotropic pathway, and a subsequent increase in the intracellular $Ca^{2+}$ concentration. Expression of Or83b alone leads to functional ion channels not directly responding to odorants, but being directly activated by intracellular cAMP or cGMP. Insect odorant receptors thus form ligand-gated channels as well as complexes of odorant-sensing units and cyclic-nucleotide-activated non-selective cation channels. Thereby, they provide rapid and transient as well as sensitive and prolonged odorant signalling.**

To investigate the mechanism of odorant receptor activation, we performed electrophysiological and fluorescence optical experiments on heterologously expressed odorant receptors[10,11] of the fruitfly (*Drosophila melanogaster*). Co-expression of Or22a and Or83b in human embryonic kidney (HEK293) cells (Supplementary Fig. 1) and stimulation with the odorant ethyl butyrate[12] led to an increased intracellular $Ca^{2+}$ concentration ($[Ca^{2+}]_i$), as previously reported[5]. The rise in the $[Ca^{2+}]_i$ level was dependent on the ethyl butyrate concentration and relied on the presence of extracellular $Ca^{2+}$, suggesting a $Ca^{2+}$ influx rather than a $Ca^{2+}$ release from intracellular stores (Supplementary Fig. 2).

Under voltage-clamp control at −60 mV in the whole-cell patch-clamp configuration, an inward current developed on short ethyl butyrate application (Fig. 1a), whereas no current was observed when HEK293 cells were transfected with either Or22a or Or83b alone ($n = 6$). The odour-induced current is carried by cations, because it remained unchanged in $Cl^-$-free solutions ($n = 5$, not shown), and is composed of an ionotropic ($I_i$) and a metabotropic ($I_m$) component. $I_i$, which does not rely on the presence of internal ATP and GTP (Fig. 1a, bottom), activates rapidly and shows a fast decay (peak at 1 s, termination at 10 s; see Supplementary Table 1). $I_m$, which relies on internal ATP and GTP (Fig. 1a, top), develops and decays considerably slower than $I_i$ (latency of 10 s, peak at 60 s, termination at 80 s; Supplementary Table 1), but is considerably larger than $I_i$ (Fig. 1c, top). Experiments with outside-out patches confirmed the whole-cell recording results (Fig. 1b, d and Supplementary Table 2). The nature of the $I_i$ component is at present unclear, but could result from a conformational change in the odorant receptor complex in response to odorant binding.

Further experiments were focused on the metabotropic current response. The ethyl-butyrate-induced currents showed no marked voltage dependence, except for a weak outward rectification (Fig. 1e, f). The instantaneous current responses following steps in the membrane potential suggest that the odorant-induced channels are constitutively active. No current was detected using symmetrical $Cs^+$ solutions (140 mM in pipette and bath solution; data not shown, $n = 6$). The current size depended on odorant concentration, saturating at about 100 nM ethyl butyrate (Fig. 1f, top), whereas the half-maximal effect was obtained at around 700 pM (Fig. 2e). Part of the receptors seem to be active even in the absence of ethyl butyrate, because the Or22a-inhibitor benzaldehyde[13] reduced the current recorded under control conditions (Fig. 1f, bottom, $n = 6$).

To test which G proteins are involved in metabotropic odorant signalling, we co-expressed Or22a with ion channels reporting G-protein activation[14]. Odorant receptors either couple to adenylyl cyclase stimulating $G_s$ proteins, thereby leading to increased cAMP levels, or to phospholipase C activating $G_q$ proteins, leading to inositol-1,4,5-trisphosphate formation[15]. To test putative $G_q$ protein coupling, we co-expressed human KCNQ4 (potassium voltage-gated channel, KQT-like subfamily, member 4) channels, which are inhibited on activation of endogenous $G_q$ proteins[14]. Application of ethyl butyrate at various concentrations had no effect on the human KCNQ4 current, although ethyl butyrate produced olfactory receptor currents (Supplementary Fig. 3). This indicates that odorant receptor activation by ethyl butyrate does not lead to $G_q$ protein activation. To test putative $G_s$ protein coupling, we first co-expressed human HCN2 (hyperpolarization-activated cyclic-nucleotide-gated potassium channel 2) channels, which report changes in cyclic nucleotide concentrations[16]. Increasing intracellular cAMP concentration due to $G_s$ protein activation causes a depolarizing shift of the steady-state activation curve and an acceleration of activation kinetics[16]. These modifications of HCN2 current characteristics were observed after ethyl butyrate stimulation when these channels were co-expressed with Or22a plus Or83b, with Or22a alone, but not with Or83b alone (Supplementary Fig. 4). Second, we co-expressed the

[1]Department of Evolutionary Neuroethology, Max Planck Institute for Chemical Ecology, Hans-Knöll-St 8, D-07745 Jena, Germany. [2]Center for Molecular Biomedicine, Institute of Molecular Cell Biology, Friedrich Schiller University, Jena, Nonnenplan 2, D-07743 Jena, Germany. [3]Center for Molecular Biomedicine, Department of Biophysics, Friedrich Schiller University Jena, Hans-Knöll-St. 2, D-07745 Jena, Germany.

cAMP-sensitive CNGA2 type of human cyclic-nucleotide-gated (CNG) channels[17]. In cells expressing Or22a plus Or83b or expressing Or22a alone, ethyl butyrate caused a concentration-dependent rise in CNGA2 current (Fig. 2a–c). Cells solely expressing CNGA2 did not respond to ethyl butyrate, but did respond to the adenylyl cyclase activator forskolin (Fig. 2b). Because CNG channels are also permeable to $Ca^{2+}$ (ref. 18), we visualized the stimulation of CNGA2 channels by $Ca^{2+}$ imaging. Ethyl butyrate application produced an increase in $[Ca^{2+}]_i$ that was most prominent near the plasma membrane (Fig. 2d). These data suggest that odorant stimulation of the 'true' olfactory receptor Or22a increases the intracellular cAMP concentration by means of a $G_s$ pathway. Finally, a radioreceptor assay revealed a significantly increased cAMP level in cells expressing Or22a and Or83b on ethyl butyrate stimulation (from 1.28 ± 0.06 pmol per mg protein (control) to 1.65 ± 0.10 pmol per mg protein (100 nM ethyl butyrate); $n = 3$ experiments; $P < 0.05$; Student's $t$-test). The ethyl-butyrate -concentration dependence of cAMP production reported by CNGA2 currents (Fig. 2c) was similar to that of receptor current stimulation (Figs 1f and 2e).

Because the G-protein-mediated signals are considerably slower than those arising from direct receptor activation (Fig. 1a), we asked whether the involvement of G proteins enhances the odour sensitivity of the signalling system. We therefore tested whether the non-hydrolysable inhibitor of G proteins guanosine 5′-(β-thio)diphosphate (GDP-β-S), intracellularly applied by means of the patch pipette, may reduce the ethyl butyrate sensitivity. Compared with the ethyl butyrate response under control conditions (effector concentration for half-maximum response, $EC_{50} = 677$ pM; Hill

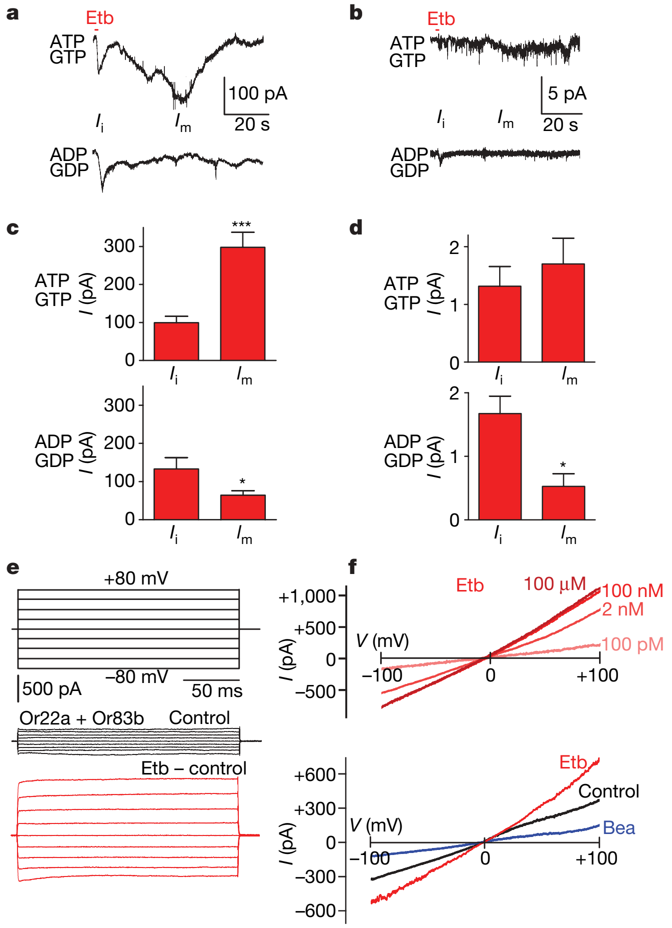



**Figure 1 | Co-expressed Or22a and Or83b mediate ethyl-butyrate-stimulated ion currents.** **a**, **c**, Recordings (**a**) and maxima (**c**) of currents activated by ethyl butyrate (Etb; 100 nM, applied for 1 s, red bar) under whole-cell voltage-clamp at –60 mV with 2 mM internal ATP and GTP (top) or ADP and GDP (bottom). Etb elicits a rapidly developing, transient inward current ($I_i$), which overlaps with a slowly activating and slowly decaying current that requires the presence of ATP and GTP ($I_m$). $n = 18$; $*P < 0.05$; $***P < 0.001$; Student's $t$-test (**c**). **b**, **d**, Recordings (**b**) and maxima (**d**) of currents activated as in **a** from outside-out patches. Without ATP and GTP, there is only the initial response of $I_i$ (bottom), whereas, in the presence of ATP and GTP, both $I_i$ and $I_m$ are activated (top). $n = 15$; $*P < 0.05$; Student's $t$-test (**d**). **e**, Current responses to the indicated pulse protocol obtained in the absence (Control) and in the presence (control currents subtracted; Etb − control) of 100 µM Etb for 1 min. **f**, Responses to ramp voltage protocols (±100 mV in 400 ms) for Etb-induced currents (at indicated Etb concentrations, control currents subtracted, top) and for currents recorded before (Control) and after application of 10 µM benzaldehyde (Bea) or 100 nM Etb (bottom). Error bars represent s.e.m.

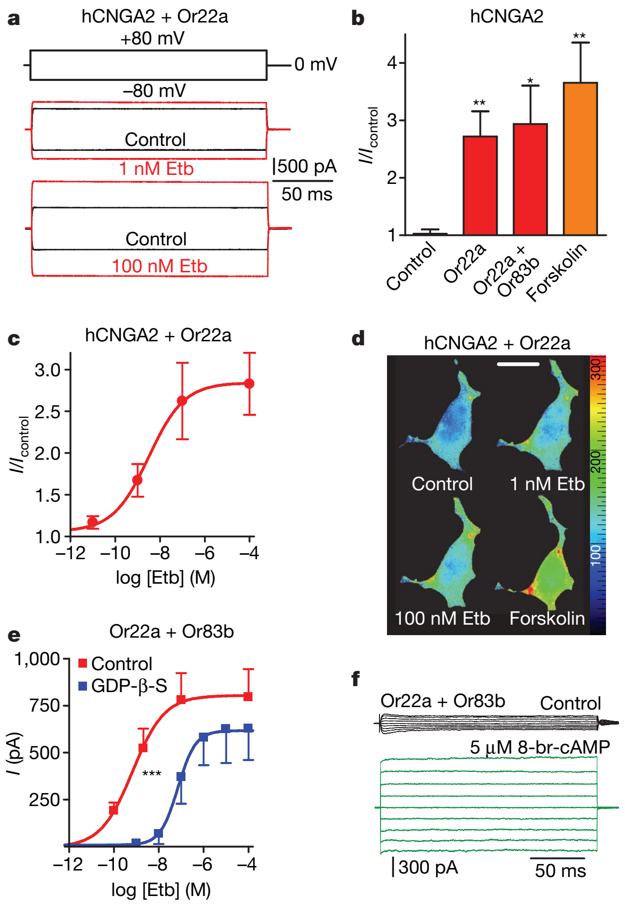



**Figure 2 | Odour signal transduction by means of $G_s$ protein signalling.** **a**, Whole-cell current responses in a cell co-expressing the cyclic-nucleotide-gated ion channel human CNGA2 (hCNGA2) with Or22a before and after application of 1 nM and 100 nM Etb. **b**, Relative increase in CNGA2 current at –80 mV induced by 100 nM Etb when hCNGA2 was expressed alone (Control) or co-expressed with the indicated odorant receptors. The orange bar indicates the stimulating effect of 10 µM forskolin in cells expressing hCNGA2 alone. $n = 5$; $*P < 0.05$; $**P < 0.01$; Student's $t$-test. **c**, Concentration–response curve for Etb-induced currents in cells expressing hCNGA2 and Or22a, measured at −80 mV and normalized to controls ($n = 5$). The continuous curve is a Hill fit ($EC_{50} = 3$ nM, Hill coefficient = 0.54). **d**, Free $[Ca^{2+}]_i$ (in nM) in a HEK293 cell expressing hCNGA2 and Or22a before (Control) and 5 min after application of Etb or 10 µM forskolin. Scale bar, 10 µm. **e**, Concentration–response curves for Etb-induced currents under control conditions and with 500 µM GDP-β-S in the pipette to disrupt G protein activation ($n = 9$, Control; $n = 11$, GDP-β-S; $***P < 0.001$; ANOVA). The continuous curves are Hill fits described by $EC_{50}$ values of 677 pM and 70 nM and Hill coefficients of 0.61 and 1.1 for the control and GDP-β-S, respectively. **f**, Current families according to the pulse protocol of Fig. 1e before (Control) and after bath application of 8-bromo-cAMP (8-br-cAMP) in a cell expressing Or83b and Or22a. Error bars represent s.e.m.

coefficient of 0.61), the presence of GDP-β-S caused a considerable shift in the ethyl-butyrate-concentration dependence ($EC_{50}$ = 70 nM; Hill coefficient of 1.1; Fig. 2e). This indicates that the ionotropic pathway of receptor activation requires significantly higher odour concentrations than the metabotropic pathway.

The cAMP signal originating from Or22a stimulation may subsequently initiate the activation of an ion channel complex requiring Or83b. If this were true, an artificial increase of the cAMP level in HEK293 cells expressing Or22a and Or83b should mimic the effect of ethyl butyrate. Indeed, bath application of the membrane-permeable cAMP analogue 8-bromo-cAMP induced a membrane current similar to that produced by ethyl butyrate (Fig. 2f). Moreover, 8-bromo-cAMP induced a similar current in cells expressing Or83b alone (Fig. 3a). The 8-bromo-cAMP failed, however, to elicit a current response in cells solely expressing Or22a (Fig. 3a). Stimulation of cAMP production by forskolin yielded large current responses of twice the size as those obtained with 8-bromo-cAMP, supporting the notion that cAMP serves as an activating ligand for Or83b (Fig. 3a). To illustrate that the current activated by 8-bromo-cAMP is identical to the current elicited by ethyl butyrate,

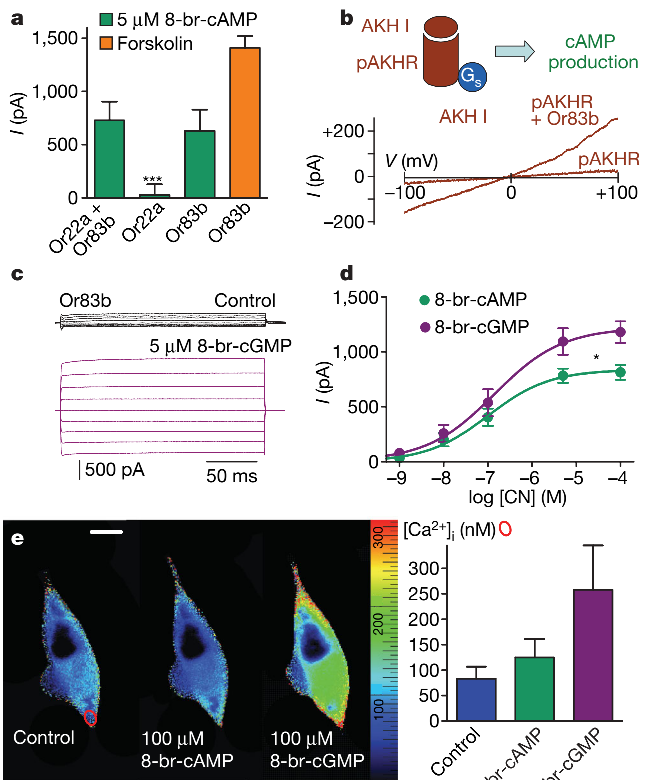



**Figure 3 | Cyclic nucleotides activate odorant receptors. a,** Maximal inward currents at –100 mV activated by 5 µM 8-bromo-cAMP (8-br-cAMP) or 50 µM forskolin in cells expressing odorant receptors as indicated ($n = 7$; ***$P < 0.001$; Student's $t$-test). **b,** Scheme of endogeneous cAMP production by AKH I-induced stimulation of the $G_s$-protein-activating *P. americana* adipokinetic hormone receptor AKHR (top). Current responses to ramp voltage protocols (±100 mV in 400 ms) recorded from cells expressing *P. americana* AKHR either alone or together with Or83b on application of 10 nM AKH I (difference curves: 5 min AKH I – before AKH I, bottom). The bath solution contained the PLC inhibitor U73122 (5 µM) to depress $G_q$-protein-related signalling. **c,** Current families recorded in a cell expressing Or83b before (Control) and after stimulation by 8-bromo-cGMP. **d,** Concentration–response curves for current activation by 8-bromo-cAMP and 8-bromo-cGMP in cells expressing Or83b (−100 mV; *$P < 0.05$; ANOVA test). **e,** Free $[Ca^{2+}]_i$ (in nM) in a cell expressing Or83b under control conditions and 3 min after application of 8-bromo-cAMP and 8-bromo-cGMP (after 5 min of washing). The histogram shows $[Ca^{2+}]_i$ in a region at the lower pole of the cell (red area). Error bars represent s.e.m.

cells expressing Or22a and Or83b were first stimulated with 8-bromo-cAMP at a high concentration of 100 µM and subsequently treated with 100 µM ethyl butyrate. Ethyl butyrate failed to cause a further increase in the current under these conditions (+1.6 ± 3.8% of current induced by 8-bromo-cAMP). To test further the notion that increased intracellular cAMP levels lead to activation of Or83b, we next co-expressed Or83b with the *Periplaneta americana* adipokinetic hormone receptor (AKHR), which is known to activate $G_s$ proteins[14]. Stimulation of *P. americana* AKHR with the adipokinetic hormone I (AKH I) induced a current (294 ± 64 pA at −100 mV, $n = 10$), but failed to do so when only AKHR was expressed (12 ± 24 pA, $n = 6$; Fig. 3b).

The results reported so far for the metabotropic odour response indicate that the *Drosophila* olfactory receptor proteins Or22a and Or83b form a complex serving as both a GPCR and an ion channel. To clarify the role of cAMP in channel activation, we tested whether the activity of a cAMP-dependent protein kinase (protein kinase A) was necessary. Ion currents elicited by 8-bromo-cAMP stimulation of Or83b-expressing cells were not significantly altered by preincubation with the protein kinase A inhibitors KT5720 (10 µM) or PKI1422 (1 µM; $n = 6$; data not shown). Thus, channel phosphorylation by protein kinase A can be excluded as an activation mechanism.

Ion channels activated or modulated by cAMP, such as CNG or HCN channels, are also sensitive to cGMP[18]. In Or83b-expressing HEK293 cells, 8-bromo-cGMP elicited a current similar to that obtained on 8-bromo-cAMP application (Fig. 3c). The concentration–response relationships for both cyclic nucleotides show that 8-bromo-cGMP is a significantly better agonist than 8-bromo-cAMP for current activation (Fig. 3d) and for the elevation of $[Ca^{2+}]_i$ (Fig. 3e).

The ion selectivity of the channels formed by Or83b was assayed by changing the composition of the bath solutions. By measuring the reversal potentials ($V_{rev}$) in $Ca^{2+}$-free and $Na^+$-free bath solutions, permeability ratios were determined according to equations (1) and (2), respectively (see Methods Summary): $P_{Na}:P_K = 1.1 \pm 0.1$ from $V_{rev}$ of +4.0 ± 1.1 mV ($Ca^{2+}$-free, $n = 7$), and $P_{Ca}:P_K = 5.4 \pm 0.2$ from $V_{rev}$ of +1.1 ± 0.4 mV ($Na^+$-free, $n = 7$). Or83b therefore forms a non-selective cation channel characterized by the permeability sequence $P_{Ca}:P_{Na}:P_K = 5.4:1.1:1$.

To test whether cyclic nucleotides also activate olfactory receptors containing odorant receptor proteins different from Or22a, Or83b was co-expressed with Or47a. Cells responding to the Or47a agonist pentyl acetate[19] were also responsive to 8-bromo-cAMP and 8-bromo-cGMP, indicating that odorant receptor complex activation by cyclic nucleotides is not a phenomenon restricted to Or22a (Supplementary Fig. 5).

Comparing the results for activation and ion selectivity of Or83b channels with those of CNG and HCN channels (both non-selective cation channels that pass monovalent cations) reveals similarities but also differences. The activation of Or83b is similar to that of CNG channels. Furthermore, $Ca^{2+}$ ions permeate Or83b and CNG channels[20] much more readily than they permeate HCN channels[18]. However, although $Cs^+$ easily permeates CNG channels, it blocks Or83b moderately and HCN potently[18].

In green algae, two GPCRs form light-gated ion channels: the proton-conducting channelrhodopsin-1 (ref. 21) and the non-selective-cation-conducting channelrhodopsin-2 (ref. 22). Or83b shows only a low degree of sequence identity with these proteins (13.0% amino acid identity with channelrhodopsin-2). Sequence comparison of Or83b with *Drosophila* ion channels, including CNG and HCN, also reveals a low degree of similarity. However, the motif TVVGYLG (393–399), located near the cytoplasmic border of the sixth predicted transmembrane helix of Or83b, is reminiscent of the selectivity filter motif TVGYG in the pores of $K^+$ channels[23]. Considering this motif as part of a putative selectivity filter, its modification might affect the ion permeability. Deletion of two residues in this motif, as indicated in Fig. 4a ('GYG mutation'), altered the slight outward rectification of wild-type Or83b currents (Figs 1f, 2f and 3b)

into a moderate inward rectification (Fig. 4a). Using both $Ca^{2+}$-free and $Na^+$-free bath solutions, the ratio of currents measured at +100 mV and −100 mV was significantly reduced in GYG-expressing cells, indicating an attenuated $K^+$ conductance (Fig. 4b). Additionally, the GYG mutation caused a shift in the reversal potentials. $V_{rev}$ was $+24.8 \pm 2.4$ mV for $Ca^{2+}$-free ($n = 5$) and $+16.5 \pm 5.1$ mV for $Na^+$-free ($n = 7$) bath solutions (Fig. 4c). The permeability ratios determined according to equations (1) and (2) were $P_{Na}$:$P_K = 2.5 \pm 0.3$ and $P_{Ca}$:$P_K = 19.6 \pm 8.9$, indicating a reduction in $K^+$ permeability with respect to the wild type. These experiments demonstrate that Or83b forms ion channels in HEK293 cells and that the TVVGYLG motif is involved in controlling the ion permeability of the channel (Fig. 4d).

The pair of *Drosophila* olfactory receptor proteins Or22a–Or83b represents a typical insect odorant receptor. Homologues of Or83b have been identified in various other insect species. The design principle of insect odorant receptors being composed of both an odorant-sensitive protein and a protein finally transducing the chemical message into an electrical signal ensures very rapid recognition of high odour concentrations by means of the ionotropic pathway as well as a somewhat slower but prolonged and highly sensitive odour detection via the G-protein-mediated signal amplification (Fig. 4d).

***Note added in proof:*** A study by Sato *et al.*[24] shows that heteromeric insect olfactory receptor complexes from different species form ligand-gated ion channels when expressed in various heterologous systems.

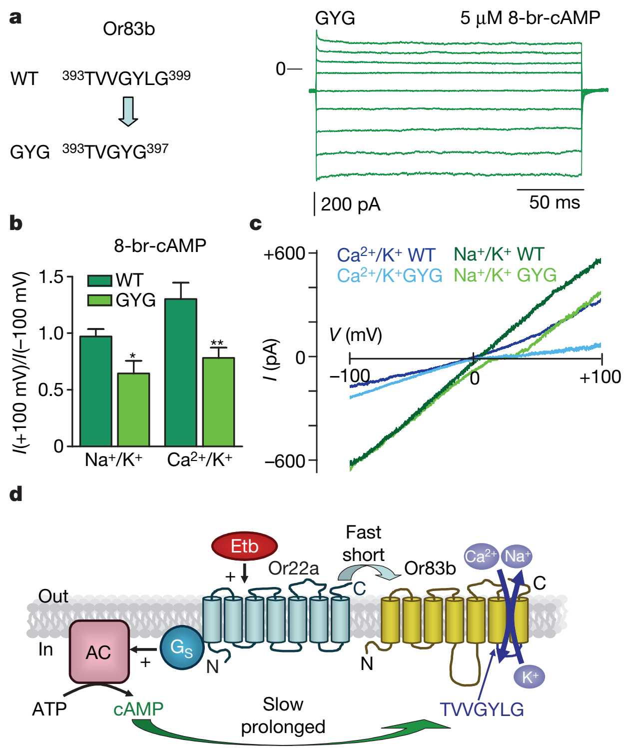



**Figure 4 | Modification of a putative Or83b pore region alters ion permeability.** **a**, The wild type (WT) Or83b amino acids Val 394 and Leu 398 were deleted in a mutant termed 'GYG' (left panel). Family of currents activated by 8-bromo-cAMP in a cell expressing the Or83b GYG mutant (right panel; recording protocol as in Fig. 1e; '0' indicates the baseline). **b**, Degree of current rectification of 8-bromo-cAMP-activated (5 µM) wild-type and mutant channels, determined as the absolute value of currents obtained at +100 mV relative to those at –100 mV. The bath/pipette solutions contained in mM: 140 $Na^+$/140 $K^+$ ($Na^+$/$K^+$) and 10 $Ca^{2+}$/140 $K^+$ ($Ca^{2+}$/$K^+$). $n = 7$; *$P < 0.05$, **$P < 0.01$; Student's *t*-test. **c**, Superimposed current responses to ramp voltage protocols (±100 mV in 400 ms) on stimulation of HEK293 cells expressing wild type or GYG mutant with 5 µM 8-bromo-cAMP under conditions as in **b**. **d**, Scheme of odorant receptor and channel function. When stimulated by an odorant such as Etb, Or22a activates the channel protein Or83b by an ionotropic and a metabotropic pathway. The direct activation of Or83b by Or22a (light blue arrow) leads to a fast and transient ion conductance. The metabotropic pathway includes activation of $G_s$ proteins by Or22a. The $G\alpha_s$ subunit stimulates adenylyl cyclase (AC), thereby increasing cAMP production. Cyclic AMP slowly activates a long-lasting non-selective cation conductance. Opening of the Or38b channel that harbours the TVVGYLG motif leads to $Na^+$ and $Ca^{2+}$ influx and membrane depolarization.

## METHODS SUMMARY

**Cell culture and transfection.** HEK293 cells were cultured as described[25]. Cells were transfected with 1 µg pcDNA3-dOr22a-GFP (green fluorescent protein) and/or 1 µg pcDNA3-dOr83b-GFP using Roti-Fect (Roth). The DNAs were a gift from E. Neuhaus and H. Hatt[5].

**Electrophysiology.** Ion currents in HEK293 cells were measured with the patch-clamp technology as described[25]. The permeability ratios were calculated according to

$$P_{Na}:P_K = [K^+]_i/[Na^+]_o \times \exp(V_{rev}F/RT) \quad (1)$$

and

$$P_{Ca}:P_K = [K^+]_i/4[Ca^{2+}]_o \times \exp(V_{rev}2F/RT) \quad (2)$$

where $P_{ion}$ are the permeabilities, $[ion]_i$ and $[ion]_o$ are the internal and external activity of the respective ion, $V_{rev}$ is the reversal potential[26], $F$ is the Faraday constant, $R$ is the gas constant, and $T$ is the absolute temperature. To rule out an effect on endogenous HEK293 channels[27], ethyl butyrate, 8-bromo-cAMP and 8-bromo-cGMP were first tested on non-transfected cells ($n = 5$–7).

**Calcium imaging.** Free intracellular $Ca^{2+}$ concentration ($[Ca^{2+}]_i$) was determined by Fura-2 fluorescence as described[28].

**Radioreceptor assay.** The cAMP content of cells was determined as described in Methods.

**Data presentation and statistics.** Results are given as means ± standard error of mean (s.e.m.), where $n$ represents the number of cells. The evaluation of statistical significance of differences was performed with Student's *t*-test for testing one variable and two-way ANOVA for testing two variables.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 23 July 2007; accepted 18 February 2008.**
**Published online 13 April 2008.**

1. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
2. Mombaerts, P. Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**, 707–711 (1999).
3. Ronnett, G. V. & Moon, C. G proteins and olfactory signal transduction. *Annu. Rev. Physiol.* **64**, 189–222 (2002).
4. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol.* **4**, e20 (2006).
5. Neuhaus, E. M. *et al.* Odorant receptor heterodimerization in the olfactory system of *Drosophila melanogaster*. *Nature Neurosci.* **8**, 15–17 (2005).
6. Larsson, M. C. *et al.* *Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).
7. Wistrand, M., Kall, L. & Sonnhammer, E. L. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Sci.* **15**, 509–521 (2006).
8. Rützler, M., Lu, T. & Zwiebel, L. J. Gα encoding gene family of the malaria vector mosquito *Anopheles gambiae*: expression analysis and immunolocalization of $AG_{\alpha q}$ and $AG_{\alpha o}$ in female antennae. *J. Comp. Neurol.* **499**, 533–545 (2006).
9. Martin, F., Charro, M. J. & Alcorta, E. Mutations affecting the cAMP transduction pathway modify olfaction in *Drosophila*. *J. Comp. Physiol. A* **187**, 359–370 (2001).
10. Wetzel, C. H. *et al.* Functional expression and characterization of a *Drosophila* odorant receptor in a heterologous cell system. *Proc. Natl Acad. Sci. USA* **98**, 9377–9380 (2001).
11. Katada, S., Nakagawa, T., Kataoka, H. & Touhara, K. Odorant response assays for a heterologously expressed olfactory receptor. *Biochem. Biophys. Res. Commun.* **305**, 964–969 (2003).

12. Dobritsa, A. A., van der Goes van Naters, W., Warr, C. G., Steinbrecht, R. A. & Carlson, J. R. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37,** 827–841 (2003).
13. Pelz, D., Roeske, T., Syed, Z., de Bruyne, M. & Galizia, C. G. The molecular receptive range of an olfactory receptor *in vivo* (*Drosophila melanogaster* Or22a). *J. Neurobiol.* **66,** 1544–1563 (2006).
14. Wicher, D. *et al.* Differential receptor activation by cockroach adipokinetic hormones produces differential effects on ion currents, neuronal activity and locomotion. *J. Neurophysiol.* **95,** 2314–2325 (2006).
15. Barry, P. H. The relative contributions of cAMP and $InsP_3$ pathways to olfactory responses in vertebrate olfactory receptor neurons and the specificity of odorants for both pathways. *J. Gen. Physiol.* **122,** 247–250 (2003).
16. Ludwig, A. *et al.* Two pacemaker channels from human heart with profoundly different activation kinetics. *EMBO J.* **18,** 2323–2329 (1999).
17. Dhallan, R. S., Yau, K. W., Schrader, K. A. & Reed, R. R. Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. *Nature* **347,** 184–187 (1990).
18. Craven, K. B. & Zagotta, W. N. CNG and HCN channels: two peas, one pod. *Annu. Rev. Physiol.* **68,** 375–401 (2006).
19. Hallem, E. A., Ho, M. G. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117,** 965–979 (2004).
20. Kaupp, U. B. & Seifert, R. Cyclic nucleotide-gated ion channels. *Physiol. Rev.* **82,** 769–824 (2002).
21. Nagel, G. *et al.* Channelrhodopsin-1: a light-gated proton channel in green algae. *Science* **296,** 2395–2398 (2002).
22. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100,** 13940–13945 (2003).
23. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of $K^+$ conduction and selectivity. *Science* **280,** 69–77 (1998).
24. Sato, K. *et al.* Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* doi:10.1038/nature06850 (this issue).
25. Wicher, D., Agricola, H. J., Schönherr, R., Heinemann, S. H. & Derst, C. TRP$\gamma$ channels are inhibited by cAMP and contribute to pacemaking in neurosecretory insect neurons. *J. Biol. Chem.* **281,** 3227–3236 (2006).
26. Hille, B. *Ion Channels of Excitable Membranes* 441–470 (Sinauer Associates, Sunderland, 2001).
27. Bugaj, V. *et al.* Functional properties of endogenous receptor- and store-operated calcium influx channels in HEK293 cells. *J. Biol. Chem.* **280,** 16790–16797 (2005).
28. Messutat, S., Heine, M. & Wicher, D. Calcium-induced calcium release in neurosecretory insect neurons: fast and slow responses. *Cell Calcium* **30,** 199–211 (2001).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank S. Arend, S. Dietel, S. Kaltofen, A. Roßner and R. Stieber for technical assistance, E. Grosse-Wilde for discussion, E. Neuhaus and H. Hatt for providing the *Or22a* and *Or83b* DNA, and M. Biel for providing human *HCN2* DNA. This study was supported by the Max Planck Society (D.W., M.C.S. and B.S.H.) and SFB 604 (TP 4, S.H.H.).

**Author Contributions** D.W. designed the experiments, performed the whole-cell patch-clamp and the calcium-imaging experiments, and analysed the data. R.B. and R.S. made the laser scanning recordings and contributed to the electrophysiological experiments. R.H. performed the cAMP radioreceptor assay. M.C.S. contributed to the bioinformatics and molecular biology effort. D.W. and S.H.H. wrote the paper. B.S.H. and M.C.S. contributed to the manuscript. All authors discussed the concepts and results, and commented on the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.W. (dwicher@ice.mpg.de).

## METHODS

**Cell culture and transfection.** HEK293 cells were cultured at a density of $\sim 2 \times 10^4$ per 35-mm dish and transfected with 1 μg pcDNA3-dOr22a-GFP[5] and 1 μg pcDNA3-dOr83b-GFP[5] using Roti-Fect transfection kit (Roth). In some experiments, we transfected cells with pcDNA3-dOr22a-GFP or pcDNA3-dOr83b-GFP alone. We also examined the properties of dOr47a, which was cloned, using standard molecular biology methods, into a pcDNA3.1(−) expression vector. To test for putative G-protein coupling, cells were transfected with the respective odorant receptor DNA together with 1.5 μg hHCN2/pcDNA3 (provided by M. Biel), with 1 μg hKCNQ4/pcDNA3.1 or with 1 μg hCNGA2/pCMV6-XL4 (Origene). For increasing the cAMP level by means of AKHR stimulation, cells were transfected with 1 μg *P. americana* AKHR/pcDNA3.1.

For electrophysiological experiments, we only used cells showing GFP fluorescence (when illuminated at 470 nm) as an indicator of putative odorant receptor expression.

In addition, we also generated amino-terminal GFP-fusion proteins of Or22a and Or83b. These were constructed according to standard molecular biology protocols. In brief, the odorant receptor genes were cloned into a pAcGFP1-C1 mammalian expression vector (Clontech Laboratories) in frame with the GFP coding sequence, with no intervening stop codon, thus fusing the *Aequorea coerulescens* (Ac) GFP molecule to the N terminus of the odorant receptor genes. These constructs were then transfected as outlined above.

Site-directed PCR mutagenesis was used to introduce two mutations into the putative pore region of Or83b. Oligonucleotide primers were designed to replace the wild-type TVVGYLG motif with TVGYG. The mutation was subsequently confirmed by means of sequencing.

**Electrophysiology.** Ion currents in HEK293 cells were measured at room temperature (23 °C) using whole-cell patch-clamp with appropriate compensation of series resistance and of capacitive currents. Additional experiments were performed in the outside-out configuration. Current measurements and data acquisition were performed using an EPC9 patch-clamp amplifier controlled by PatchMaster software (both HEKA Elektronik). Patch-clamp pipettes were fabricated from borosilicate capillaries. Pipettes for whole-cell recordings had resistances of 2–4 MΩ; for excised-patch recordings, the pipette resistance was up to 15 MΩ.

The standard pipette solution contained (in mM) 140 KCl, 4 NaCl, 2.2 $CaCl_2$, 2 Mg-ATP, 0.05 Na-GTP, 5 EGTA and 10 HEPES (pH 7.3), and the standard bath solution contained (in mM) 135 NaCl, 5 KCl, 1 $CaCl_2$, 1 $MgCl_2$, 10 HEPES and 10 glucose (pH 7.4). To differentiate between ionotropic and metabotropic pathways of receptor activation (Fig. 1a–d), the pipette solution contained either 2 mM ATP plus 2 mM GTP or 2 mM ADP plus 2 mM GDP (both for the whole-cell and the outside-out configuration). For the preparation of $Cl^-$-free solutions, the chloride salts were substituted by gluconate salts. The pipette $Cs^+$ solution contained (in mM) 140 CsCl, 1 Mg-ATP, 10 EGTA and 10 HEPES (pH 7.3), and the $Cs^+$ bath solution contained (in mM) 140 CsCl, 15 glucose and 15 HEPES (pH 7.4). The $Ca^{2+}$-free bath solution contained (in mM) 140 NaCl, 15 glucose and 15 HEPES, and the $Na^+$-free bath solution contained (in mM) 140 *N*-methyl-D-glucamine (NMDG)-Cl, 10 $CaCl_2$, 10 glucose and 15 HEPES.

**Calcium imaging.** Cells were loaded with Fura-2 by incubation in bath solution containing 2 μM Fura-2/acetomethylester for 20 min. Free intracellular $Ca^{2+}$ concentration ($[Ca^{2+}]_i$) was determined using the fluorescence ratio method. Excitation of Fura-2 at 340 nm and 380 nm was performed with a monochromator (Polychrome II, TILL Photonics) coupled by means of an epifluorescence condenser into an Axioskop FS microscope (Carl Zeiss) with a water immersion objective (LUMPFL 40xW/IR/0.8; Olympus). Emitted light was separated by a 400-nm dichroic mirror and filtered with a 420-nm long-pass filter.

$[Ca^{2+}]_i$ was calculated according to the equation

$$[Ca^{2+}]_i = K_{eff}(R - R_{min})/(R_{max} - R) \quad (3)$$

$K_{eff}$, $R_{min}$ and $R_{max}$ were determined using HEK293 cells permeabilized with 2 μM ionomycin and three solutions with different concentrations of free $Ca^{2+}$ ($Ca^{2+}$-free, 5 mM $Ca^{2+}$ and 500 nM $Ca^{2+}$; the composition of the 500 nM solution was calculated with WEBMAXC v.2.20)[29]. The values of $K_{eff}$, $R_{min}$ and $R_{max}$ were 3.72 μM, 0.38 μM and 5.9 μM, respectively.

Fluorescence images were acquired using a cooled CCD camera controlled by TILLVision 4.0 software (TILL Photonics). The resolution was 640 × 480 pixel in a frame of 175 × 130 μm (×40/IR/0.8 objective). Image pairs were obtained by excitation for 100 ms at 340 nm and 380 nm; background fluorescence was subtracted.

**Radioreceptor assay.** Monolayers of HEK283 cells were incubated for 30 min in HEPES Ringer's buffer containing 1 mM isobutylmethylxanthine, and stimulated with 100 nM ethyl butyrate for 15 min. The reaction was stopped with 96% ethanol, and after evaporation, 50 mM Tris/4 mM EDTA (pH 7.5) was added. The cAMP content of the cellular extracts was measured by a cAMP [$^3$H] assay system (GE Healthcare) using the affinity of a purified cAMP-binding protein and a charcoal separation step according to the instructions of the manufacturer. Proteins were measured in parallel samples, and intracellular cAMP concentration was expressed as pmol $mg^{-1}$ cell protein.

**Substance application.** With the exception of GDP-β-S, which was applied by means of the patch pipette, all substances were applied to the bath using either a bath perfusion system (BPS4 from ALA) or a rapid solution changer (RSC160 from Biologic), which were controlled by the PatchMaster software (HEKA Elektronik).

**Data analysis.** For data analysis, the software IgorPro (WaveMetrics) or Prism 4 (Graph Pad Software) were used.

**Chemicals.** 8-bromo-cAMP, 8-bromo-cGMP, benzaldehyde, ethyl butyrate, forskolin, GDP-β-S, KT5720 and pentyl acetate were obtained from Sigma; myristoylated protein kinase A inhibiting peptide 14–22 amide (PKI1422) and U73122 were from Calbiochem; and Fura-2 and wheat germ agglutin were from Invitrogen.

29. Patton, C., Thompson, S. & Epel, D. Some precautions in using chelators to buffer metals in biological solutions. *Cell Calcium* 35, 427–431 (2004).

# LETTERS

# Upper intestinal lipids trigger a gut–brain–liver axis to regulate glucose production

Penny Y. T. Wang[1], Liora Caspi[1], Carol K. L. Lam[1,2], Madhu Chari[1,2], Xiaosong Li[4], Peter E. Light[5], Roger Gutierrez-Juarez[4], Michelle Ang[1], Gary J. Schwartz[4] & Tony K. T. Lam[1,2,3]

**Energy and glucose homeostasis are regulated by food intake and liver glucose production, respectively. The upper intestine has a critical role in nutrient digestion and absorption. However, studies indicate that upper intestinal lipids inhibit food intake as well in rodents and humans by the activation of an intestine–brain axis[1–4]. In parallel, a brain–liver axis has recently been proposed to detect blood lipids to inhibit glucose production in rodents[5]. Thus, we tested the hypothesis that upper intestinal lipids activate an intestine–brain–liver neural axis to regulate glucose homeostasis. Here we demonstrate that direct administration of lipids into the upper intestine increased upper intestinal long-chain fatty acyl-coenzyme A (LCFA-CoA) levels and suppressed glucose production. Co-infusion of the acyl-CoA synthase inhibitor triacsin C or the anaesthetic tetracaine with duodenal lipids abolished the inhibition of glucose production, indicating that upper intestinal LCFA-CoAs regulate glucose production in the preabsorptive state. Subdiaphragmatic vagotomy or gut vagal deafferentation interrupts the neural connection between the gut and the brain, and blocks the ability of upper intestinal lipids to inhibit glucose production. Direct administration of the *N*-methyl-D-aspartate ion channel blocker MK-801 into the fourth ventricle or the nucleus of the solitary tract where gut sensory fibres terminate abolished the upper-intestinal-lipid-induced inhibition of glucose production. Finally, hepatic vagotomy negated the inhibitory effects of upper intestinal lipids on glucose production. These findings indicate that upper intestinal lipids activate an intestine–brain–liver neural axis to inhibit glucose production, and thereby reveal a previously unappreciated pathway that regulates glucose homeostasis.**

Intraduodenal administration of lipids, in particular long-chain fatty acids (LCFA), rapidly inhibits food intake in both rodents and humans[6,7]. The small intestinal lipids activate sensory (afferent) neurons that project directly to the nucleus of the solitary tract (NTS)[8]. Intraduodenal infusion of the topical anaesthetic tetracaine blocks neural communication of intestinal vagal sensory fibres and abolishes the inhibitory feeding effects of intestinal lipids[6]. Furthermore, intestinal lipids fail to reduce food intake in rodents with gut vagal deafferentation[9]. Together, these findings indicate that upper intestinal lipids negatively regulate food intake in the preabsorptive state by activation of the intestine–brain neural axis. On nutrient absorption, the brain directly detects a rise in both blood nutrients and hormones to negatively regulate food intake and maintain energy balance[10–14]. Recent studies indicate that the hypothalamic sensing mechanisms regulate peripheral metabolic processes as well[15–20]. A selective increase of hypothalamic lipids caused by either lipid administration[5] or inhibition of hypothalamic carnitine palmitoyl-transferase-1 (ref. 20) lowers glucose production (GP) and plasma glucose levels. Furthermore, hepatic vagotomy abolishes the ability of the hypothalamic lipids to lower GP[5]. These studies indicate that hypothalamic nutrients activate a brain–liver neural axis to maintain glucose homeostasis.

On the basis of these findings, we tested the hypothesis that acute accumulation of lipids in the upper intestine triggers an intestine–brain–liver neurocircuitry in the preabsorptive state to regulate glucose homeostasis. Together with findings demonstrating the role of intestinal lipids in food intake[6,7], we propose that intestinal-lipid-sensing mechanisms represent one of the first lines of metabolic defenses against nutrient excess, providing energy and metabolic balance by rapidly downregulating both GP and food intake on nutrient exposure.

To examine whether upper intestinal lipids regulate GP in the preabsorptive state (Fig. 1a), we first established an intraduodenal lipid-infused *in vivo* model. On the day of the *in vivo* experiment, the fully recovered conscious rats were first subjected to a pancreatic (basal insulin) clamp to obtain steady state basal measurements of glucose metabolism (Fig. 1b and Supplementary Table 1). Intralipid was then administered directly into the duodenum at 0.03 kcal $min^{-1}$ for 50 min (Fig. 1b). Intraduodenal administration of lipids increased upper intestinal LCFA-CoA levels (Fig. 1c) but did not alter plasma and portal free fatty acids (FFA)/triglycerides (TG) (Supplementary Fig. 1a, b) or liver LCFA-CoA levels (Supplementary Fig. 1c). These findings are consistent with previous reports indicating that intraduodenal lipid administered at 0.22 kcal $min^{-1}$ (for 30–40 min) did not affect plasma lipid levels[21]. In the presence of basal circulating lipid and insulin levels (Supplementary Fig. 1a, b and Supplementary Table 1), intraduodenal lipid administration increased the exogenous glucose infusion rate by 2.5-fold compared to the vehicle to maintain euglycemia during the pancreatic clamp (Fig. 1d and Supplementary Table 1). The increase in glucose infusion rate was fully accounted for by a ~55% inhibition of GP (Fig. 1e, f), because the rate of glucose uptake remained constant (Fig. 1g). Under non-clamped physiological settings, intraduodenal lipid administration significantly lowered plasma glucose levels by 15 min, and this glucose-lowering effect persisted throughout the 50 min protocol (Fig. 1h). Intraduodenal lipid administration did not alter peripheral and portal FFA/TG levels (Supplementary Fig. 1d) but lowered insulin levels (Fig. 1h) by 15 min (when the plasma glucose drop first became evident), indicating an enhancement of insulin sensitivity. These data collectively indicate that an acute rise in upper intestinal lipids decreased GP and plasma glucose levels.

We next co-infused the inhibitor of acyl-CoA synthetase triacsin C with lipids to investigate whether the formation of upper intestinal esterified lipids (LCFA-CoA) is required to regulate GP.

[1]Toronto General Hospital Research Institute, University Health Network, Toronto M5G 1L7, Canada. [2]Department of Physiology, and [3]Department of Medicine, University of Toronto, Toronto M5S 1A8, Canada. [4]Departments of Medicine and Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [5]Department of Pharmacology, University of Alberta, Edmonton T6G 2H7, Canada.

Co-administration of triacsin C (80 μM in 0.5 ml) with lipids abolished the increase in upper intestinal LCFA-CoA levels (Fig. 1c) and glucose infusion rate (Fig. 1d) and the reduction in GP (Fig. 1e, f) induced by lipids. Again, there was no difference in glucose uptake between groups (Fig. 1g). Administration of triacsin C alone did not affect peripheral glucose metabolism (Fig. 1d–g and Supplementary Table 2). Thus, we have demonstrated for the first time, to our knowledge, that an acute increase of upper intestinal LCFA-CoAs suppressed GP.

We propose that preabsorptive upper intestinal lipids first elicit an afferent neuronal signal to the hindbrain. The hindbrain then processes this and relays the signal to the liver to lower GP (Fig. 2a). To test this hypothesis, we first infused the topical anaesthetic tetracaine (0.5 mg in 0.5 ml) locally to inhibit the neurotransmission of vagal afferent fibres innervating the upper intestine (Fig. 2b). Intraduodenal infusion of tetracaine alone did not affect whole-body glucose metabolism (Fig. 2c–f and Supplementary Table 2). In contrast, tetracaine administration, when co-infused with lipids, significantly reduced the ability of duodenal lipids to enhance peripheral insulin sensitivity and to lower GP (Fig. 2c–e). These data indicate that preabsorptive lipids regulate GP, possibly through a neuronal network. To assess the potential role of the gut extrinsic sensory innervation in this response, we performed intraduodenal lipid infusion in rats subjected to either subdiaphragmatic vagatomy or gut vagal deafferentation (Fig. 2a, b). Both surgical procedures interrupt the vagal afferent innervations from the gut to the brain but when performed alone did not affect peripheral glucose metabolism within two weeks (Fig. 2c–f and Supplementary Table 2). In contrast, intraduodenal lipids failed to increase the glucose infusion rate and lower GP during the clamp in rats subjected to either subdiaphragmatic vagatomy or gut vagal deafferentation (Fig. 2c–e). Glucose uptake (Fig. 2f), plasma insulin and glucose levels (Supplementary Table 1) were comparable in all groups. Taken together, these independent yet complementary findings suggest that the vagal afferent innervation is required for the neurotransmission triggered by upper intestinal lipids to lower GP.

The caudal hindbrain NTS senses afferent neuronal signals to maintain energy balance[1–4]. Glutamatergic neurotransmission has been demonstrated in gut-recipient NTS neurons, and *N*-methyl-D-aspartate (NMDA) ion channels and receptors have been localized to vagal afferent terminals in the hindbrain NTS[22,23]. Importantly, direct administration of the NMDA channel blocker MK-801 into the fourth ventricle[24] or the NTS[25] extends meal duration, probably by modulating glutamatergic transmission at central vagal afferent terminals. Here we tested the hypothesis that glutamatergic transmission in the NTS relays afferent neuronal signals from the duodenal lipids to the liver to subsequently lower GP (Fig. 3a). We implanted a single catheter into the fourth ventricle and performed intracerebroventricular (ICV) administration of MK-801 into the fourth ventricle (0.42 μg min$^{-1}$ at a rate of 0.083 μl min$^{-1}$) in fully recovered rats (Fig. 3b). ICV infusion of MK-801 into the fourth ventricle began with the start of the clamp and was maintained

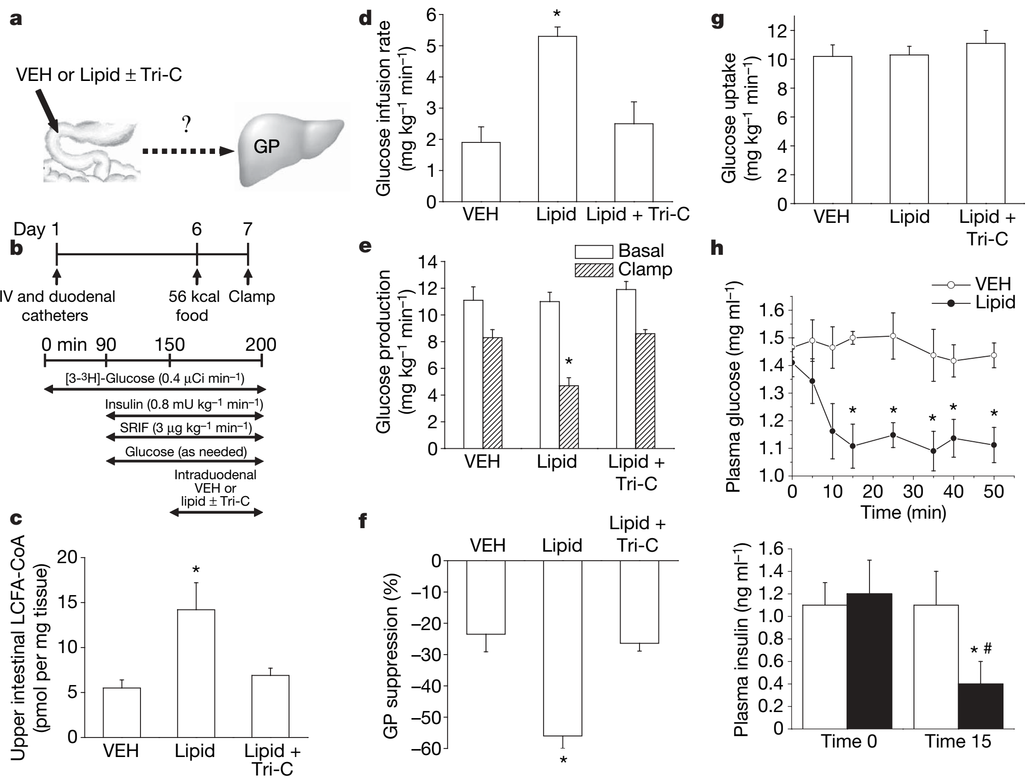



**Figure 1 | Upper intestinal lipids suppress liver glucose production.** **a**, Schematic representation of the working hypothesis. Lipid with or without triacsin C (Tri-C) was infused through a duodenal catheter. VEH, vehicle. **b**, Experimental procedure and clamp protocol. SRIF, somatostatin. Duodenal catheter or venous and arterial (IV) catheters were implanted on day 1. On the evening before the pancreatic (basal insulin) clamp studies, all rats received a fixed portion of calories (~56 kcal) to ensure a comparable post-absorptive state at the start of the clamp experiments. **c**–**e**, During the pancreatic clamp, intraduodenal lipid infusion increased upper intestinal LCFA-CoA levels (**c**, $*P < 0.05$ versus other groups), increased glucose infusion rate (**d**, $*P < 0.01$ versus other groups) and decreased GP (**e**, $*P < 0.001$ versus other groups). In contrast, co-infusion with Tri-C negated the effects of lipids on upper intestinal LCFA-CoA levels (**c**), glucose infusion rate (**d**) and GP (**e**). **f**, Suppression of GP during the clamp period expressed as the percentage decrease from basal GP ($*P < 0.001$ versus other groups). **g**, Glucose uptake was unaltered in all groups. Administration of intraduodenal Tri-C alone ($n = 4$) and saline ($n = 5$) were combined as one vehicle (VEH) group. VEH ($n = 9$), Lipid ($n = 7$), Lipid + Tri-C ($n = 6$). **h**, Under unclamped settings, intraduodenal lipid infusion lowered plasma glucose levels in 15 min ($n = 5$). This effect persisted throughout the 50 min infusion period ($*P < 0.01$ versus saline (VEH) ($n = 5$)). Intraduodenal lipid infusion for 15 min lowered plasma insulin levels ($*P < 0.01$ versus saline, $^{\#}P < 0.05$ versus pre-intraduodenal infusion). Values are shown as mean ± s.e.m.

throughout the experiment to inhibit NMDA neurotransmission in the hindbrain (Fig. 3a, b). Intraduodenal lipid administration was initiated one hour after ICV administration of MK-801 into the fourth ventricle (Fig. 3b). Administration of ICV MK-801 into the fourth ventricle alone did not affect whole-body glucose metabolism but attenuated the effects of duodenal lipids to increase the glucose infusion rate and to lower GP (Fig. 3c–f and Supplementary Table 2). To explore the neuroanatomical location of the blockage effects on duodenal lipids in the hindbrain, we implanted bilateral catheters into the NTS (Fig. 3a). We infused MK-801 at a 15-fold-lower dose than the ICV administration into the fourth ventricle on each catheter into the NTS (0.03 $\mu$g min$^{-1}$ at a rate of 0.006 $\mu$l min$^{-1}$). NTS MK-801 alone did not affect whole-body glucose metabolism (Fig. 3b–f and Supplementary Table 2). However, NTS MK-801 administration prevented the increase in glucose infusion rate and the reduction in GP induced by duodenal lipids (Fig. 3c–e). Glucose uptake (Fig. 3f), plasma glucose and insulin levels (Supplementary Table 1) were comparable in all groups. These results support the idea that NTS glutamatergic transmission relays afferent neuronal signals triggered by duodenal lipids to lower GP.

To begin evaluating whether the hindbrain relays signals induced by upper intestinal lipids to the liver to lower GP, we repeated the intraduodenal lipid infusion protocol in combination with the clamp technique in rats that underwent vagotomy of the hepatic branch. Hepatic vagotomy negated the ability of intraduodenal lipids to increase glucose infusion rate and lower GP (Fig. 3c–e), and when performed alone did not affect whole-body glucose metabolism within two weeks (Fig. 3c–f and Supplementary Table 2). Glucose uptake (Fig. 3f) was comparable in all groups. These data strengthen the working hypothesis that the upper-intestinal-lipid-induced neuronal signals are first processed by the NTS and are then relayed to the liver to lower GP.

We next studied whether the activation of the intestine–brain–liver neural axis by administration of upper intestinal lipids is impaired in the early onset of diet-induced insulin-resistant rodents. Rats fed on a lard-oil-enriched diet rapidly increased their caloric intake (Supplementary Fig. 2a, b) and acquired brain defect(s) in detecting lipids to reduce GP after three days of high caloric intake[26]. We repeated the intraduodenal lipid administration protocol in combination with the clamp technique in this early-onset high-fat diet (HFD)-induced insulin-resistant model. In comparison with rats fed on a regular chow diet, intraduodenal lipid administration failed to increase glucose infusion rate and to lower GP in HFD rats (Fig. 3g–i). These data suggest that the metabolic effects of duodenal lipid are altered, leading to dysregulation of GP in the early onset of diet-induced insulin resistance and obesity.

The body adapts to nutrient excess through multiple redundant nutrient-sensing mechanisms in various tissues. The small intestine and the brain have both been implicated in sensing nutrients and maintaining energy balance by downregulating food intake[1–4,10–14]. Here we report that an increase in the upper intestinal LCFA-CoA level suppressed GP and the plasma glucose level in the preabsorptive state in rodents. This was achieved by triggering an intestine–brain–liver neuronal network. In parallel, accumulation of LCFA-CoAs in the hypothalamus lowers GP and plasma glucose levels[27]. These data suggest that the small intestine and the brain have redundant LCFA-CoA-sensing mechanisms to maintain glucose homeostasis by inhibiting GP. The caudal brainstem NTS seems to be a critical site on

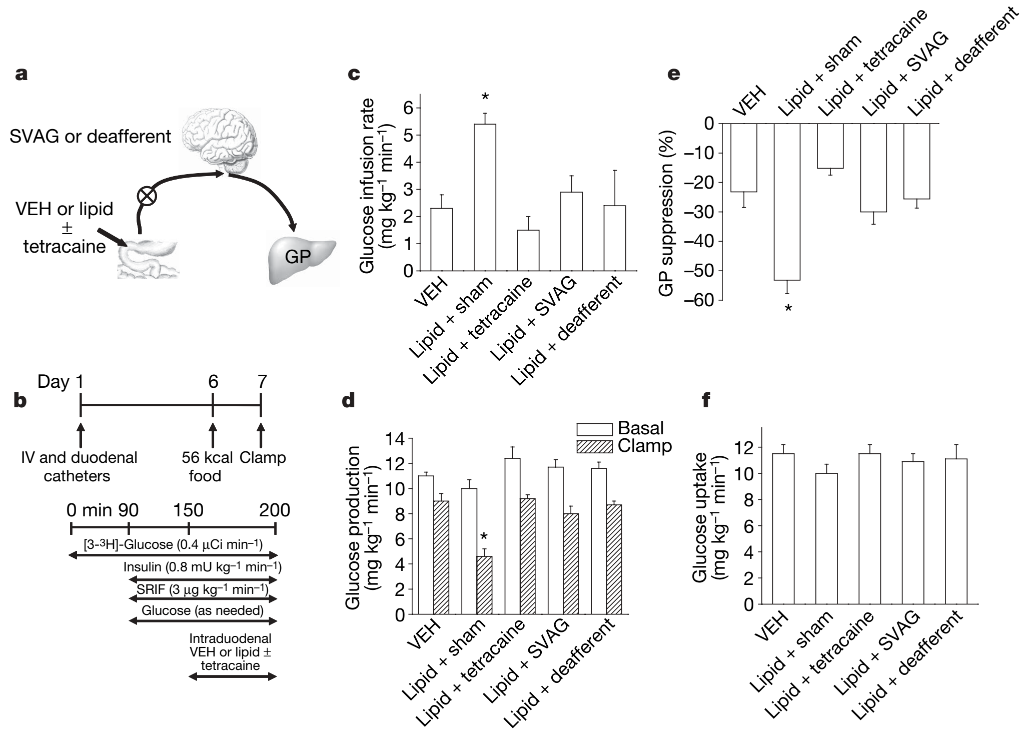



**Figure 2 | Upper intestinal lipids suppress glucose production through a neuronal network. a**, Schematic representation of the working hypothesis. Duodenal lipid was co-infused with tetracaine, which abolishes the ascending neuronal signal to the brain. In separate studies, duodenal lipid was infused into rats that underwent subdiaphragmatic vagotomy (SVAG), gut vagal deafferentation (deafferent) or sham operation. **b**, Experimental procedure and clamp protocol. SVAG, deafferent or sham operation was performed about one week prior to the implantation of the duodenal and IV catheters. Rats were restricted to ~56 kcal of food one night before the pancreatic clamp. **c**, **d**, Intraduodenal lipid infusion increased the glucose infusion rate (**c**, $*P < 0.01$ versus other groups) and lowered GP (**d**, $*P < 0.001$ versus all other groups) in normal ($n = 2$) and sham-operated ($n = 5$) rats. Rats receiving duodenal tetracaine administration, SVAG or deafferent failed to respond to upper intestinal lipids. **e**, Suppression of GP during the clamp period expressed as the percentage decrease from basal GP ($*P < 0.001$ versus other groups). **f**, Glucose uptake was unaltered in all groups. Intraduodenal tetracaine infusion alone ($n = 4$), saline-infused rats with SVAG ($n = 5$) and saline-infused rats with deafferent ($n = 4$) were combined as one VEH group. Values are shown as mean + s.e.m. VEH ($n = 13$), Lipid in normal and sham-operated rats ($n = 7$), Lipid + tetracaine ($n = 6$), Lipid + SVAG ($n = 5$), and Lipid + deafferent ($n = 5$).

which duodenal-lipid-stimulated vagal afferent signals act on to modulate GP. The ability of the NTS to mediate hypothalamic lipid signals that regulate GP, however, remains to be elucidated. The NTS could serve as a critical site of neural circuit convergence in integrating nutrient-sensing signals in the body to negatively regulate GP. Additionally, future studies are needed to examine the potential cross-talk within brain regions (that is, NTS and the arcuate nucleus) in mediating intestinal-lipid-sensing mechanisms that regulate glucose homeostasis by means of the intestine–brain–liver neural axis.

We report that the formation of upper intestinal LCFA-CoAs was required for upper intestinal lipids to lower GP. Although it still remains to be clarified whether enterocytes or enteroendocrine cells within the gut are mediating the effects of intestinal LCFA-CoA, our findings suggest that local intestinal biochemical events triggered by lipid administration are critical in mediating the lipid-induced suppression of GP. Intestinal hormones released by the presence of nutrients in the intestinal lumen provide an attractive possible set of mediators of this GP-lowering response. For example, upper intestinal lipid infusions induce local release of the gastrointestinal satiety peptide cholecystokinin (CCK)[2,7]. CCK binds to CCK-A receptors on gut vagal afferent neurons and increases neurophysiological activity in duodenal and gastric fibres that are also sensitive to two meal-related stimuli: nutrients and mechanical distension. Surgical and chemical interruption of gut vagal afferents attenuates the satiety effects of CCK and nutrient secretagogue of CCK, such as lipid[2,7]. The potential involvement of CCK in mediating the lipid-induced intestine–hindbrain–liver neurocircuitry warrants future investigation in both rodents and humans. Furthemore, our data indicate that in the early onset of diet-induced insulin resistance, the intestine may have acquired defect(s) in lipid sensing, hindering glucose homeostasis regulation. These data may begin to shed light on the potential pathogenic effects of gut microbiota[28,29] and the

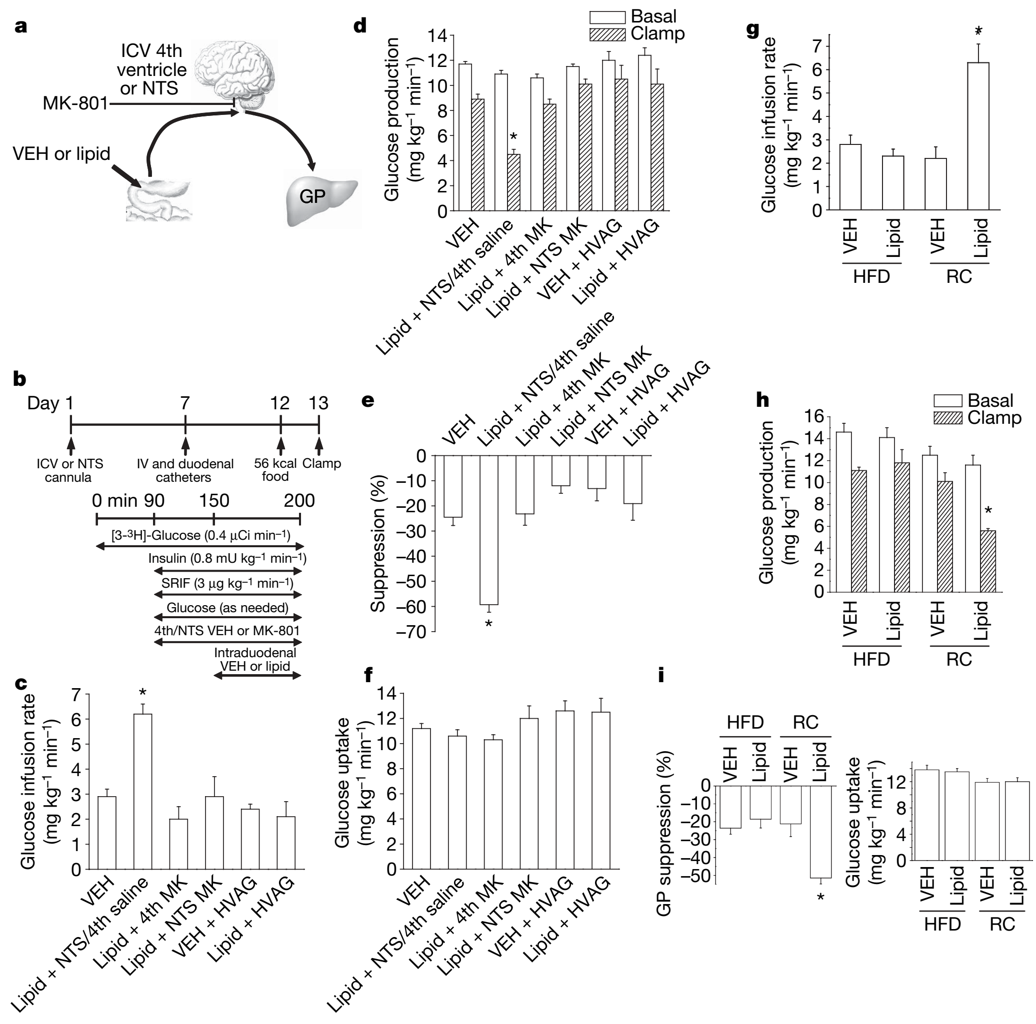



**Figure 3 | Upper intestinal lipids suppress glucose production by activating an intestine–NTS–liver neurocircuitry.** **a**, Schematic representation of the working hypothesis. NMDA channel inhibitor MK-801 was infused directly into the fourth ventricle (ICV 4th) or the NTS. **b**, Experimental procedure and clamp protocol. Stereotaxic surgeries were conducted 7 days before placing IV and duodenal catheters. All rats were restricted to ~56 kcal one night before the pancreatic-insulin clamp. **c**, **d**, Intraduodenal-lipid infusion increased the glucose infusion rate (**c**, $^*P < 0.01$ versus other groups) and lowered GP (**d**, $^*P < 0.001$ versus all other groups) in rats with ICV infusion of saline into the fourth ventricle ($n = 4$) or into the NTS ($n = 3$). Rats receiving ICV infusion of MK-801 into the fourth ventricle (4th MK), MK-801 into the NTS (NTS MK) or HVAG failed to respond to upper intestinal lipids to increase glucose infusion rate and lower GP. **e**, Suppression of GP during the clamp period expressed as the percentage decrease from basal GP ($^*P < 0.001$ versus other groups). **f**, Glucose uptake was unaltered in all groups. Intraduodenal-saline-infused rats with ICV infusion of saline into the fourth ventricle ($n = 4$), ICV infusion of MK-801 into the fourth ventricle (4th MK; $n = 4$), infusion of saline into the NTS ($n = 4$) and infusion of MK-801 into the NTS ($n = 4$) were combined as one VEH group. VEH ($n = 16$), Lipid + ICV fourth saline or Lipid + NTS saline ($n = 7$), Lipid + ICV fourth MK-801 ($n = 5$), Lipid + NTS MK-801 ($n = 5$). Saline + HVAG ($n = 5$), Lipid + HVAG ($n = 5$). **g–i**, During the pancreatic (basal insulin) clamp, rats that doubled their daily caloric intake for three days (HFD; $n = 8$) failed to respond to intraduodenal lipid infusion to increase glucose infusion rate (**g**) and lower GP (**h**, **i**) compared to rats fed on a regular chow (RC; $n = 8$). **i**, Glucose uptake was comparable in all groups. HFD saline (VEH; $n = 4$), RC saline (VEH; $n = 6$). Values are shown as mean + s.e.m.

beneficial metabolic effects of gastric bypass surgeries in obesity and diabetes[30].

The small intestine has developed lipid-sensing mechanisms to maintain energy balance by inhibiting food intake in both rodents and humans. Here we have provided the first evidence, to our knowledge, that a selective increase in upper intestinal lipids lowers GP through an intestine–brain–liver neurocircuitry in rodents. This discovery may reveal novel sites and targets to lower GP in diabetes and obesity.

## METHODS SUMMARY

Rats were subjected to cannulations in the duodenum, the internal jugular vein and the carotid artery for the pancreatic clamp *in vivo* infusion studies. Rats also received subdiaphragmatic vagotomy, gut vagal deafferentation, hepatic vagotomy, or stereotaxic surgery that implanted a single catheter into the fourth cerebral ventricle (ICV fourth) or bilateral catheters into the NTS. The *in vivo* infusion experiments lasted a total of 200 min, with pancreatic clamps (insulin at 0.8 mU $kg^{-1}$ $min^{-1}$ and somatostatin at 3 μg $kg^{-1}$ $min^{-1}$) initiated at 90 min and intraduodenal infusions began at 150 min. A different group of rats also received ICV fourth or NTS continuous administrations at 90 min. Upper intestinal and liver tissues as well as portal plasma samples were taken shortly after the clamps to assess lipids level. For the HFD studies, rats underwent duodenal/intravenous cannulations and were put on a lard-oil-enriched HFD to induce insulin resistance. After 3 days of high caloric intake, the rats underwent the pancreatic clamp studies with intraduodenal infusions. Two additional sets of intraduodenal lipid/saline infusion experiments were performed for 15 min or 50 min in unclamped settings to evaluate the impact on plasma metabolite levels. All study protocols were reviewed and approved by the Institutional Animal Care and Use Committee of the University Health Network.

**Received 15 January; accepted 22 February 2008.**
**Published online 9 April 2008.**

1. Badman, M. K. & Flier, J. S. The gut and energy balance: visceral allies in the obesity wars. *Science* **307,** 1909–1914 (2005).
2. Cummings, D. E. & Overduin, J. Gastrointestinal regulation of food intake. *J. Clin. Invest.* **117,** 13–23 (2007).
3. Moran, T. H. & Schwartz, G. J. Neurobiology of cholecystokinin. *Crit. Rev. Neurobiol.* **9,** 1–28 (1994).
4. Murphy, K. G. & Bloom, S. R. Gut hormones and the regulation of energy homeostasis. *Nature* **444,** 854–859 (2006).
5. Lam, T. K. *et al.* Hypothalamic sensing of circulating fatty acids is required for glucose homeostasis. *Nature Med.* **11,** 320–327 (2005).
6. Greenberg, D., Smith, G. P. & Gibbs, J. Intraduodenal infusions of fats elicit satiety in sham-feeding rats. *Am. J. Physiol.* **259,** R110–R118 (1990).
7. Matzinger, D. *et al.* The role of long chain fatty acids in regulating food intake and cholecystokinin release in humans. *Gut* **46,** 689–693 (2000).
8. Monnikes, H. *et al.* Pathways of Fos expression in locus ceruleus, dorsal vagal complex, and PVN in response to intestinal lipid. *Am. J. Physiol.* **273,** R2059–R2071 (1997).
9. Sclafani, A., Ackroff, K. & Schwartz, G. J. Selective effects of vagal deafferentation and celiac-superior mesenteric ganglionectomy on the reinforcing and satiating action of intestinal nutrients. *Physiol. Behav.* **78,** 285–294 (2003).
10. Minokoshi, Y. *et al.* AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428,** 569–574 (2004).
11. Lam, T. K., Schwartz, G. J. & Rossetti, L. Hypothalamic sensing of fatty acids. *Nature Neurosci.* **8,** 579–584 (2005).
12. Cota, D. *et al.* Hypothalamic mTOR signaling regulates food intake. *Science* **312,** 927–930 (2006).
13. Schwartz, M. W., Woods, S. C., Porte, D. Jr, Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404,** 661–671 (2000).
14. Coll, A. P., Farooqi, I. S. & O'Rahilly, S. The hormonal control of food intake. *Cell* **129,** 251–262 (2007).
15. Coppari, R. *et al.* The hypothalamic arcuate nucleus: A key site for mediating leptin's effects on glucose homeostasis and locomotor activity. *Cell Metab.* **1,** 63–72 (2005).
16. Bence, K. K. *et al.* Neuronal PTP1B regulates body weight, adiposity and leptin action. *Nature Med.* **12,** 917–924 (2006).
17. Lam, T. K., Gutierrez-Juarez, R., Pocai, A. & Rossetti, L. Regulation of blood glucose by hypothalamic pyruvate metabolism. *Science* **309,** 943–947 (2005).
18. Schwartz, M. W. & Porte, D. Jr. Diabetes, obesity, and the brain. *Science* **307,** 375–379 (2005).
19. Lam, T. K. *et al.* Brain glucose metabolism controls the hepatic secretion of triglyceride-rich lipoproteins. *Nature Med.* **13,** 171–180 (2007).
20. Obici, S., Feng, Z., Arduini, A., Conti, R. & Rossetti, L. Inhibition of hypothalamic carnitine palmitoyltransferase-1 decreases food intake and glucose production. *Nature Med.* **9,** 756–761 (2003).
21. Greenberg, D., Kava, R. A., Lewis, D. R., Greenwood, M. R. & Smith, G. P. Time course for entry of intestinally infused lipids into blood of rats. *Am. J. Physiol.* **269,** R432–R436 (1995).
22. Aicher, S. A., Sharma, S. & Pickel, V. M. *N*-methyl-D-aspartate receptors are present in vagal afferents and their dendritic targets in the nucleus tractus solitarius. *Neuroscience* **91,** 119–132 (1999).
23. Berthoud, H. R., Earle, T., Zheng, H., Patterson, L. M. & Phifer, C. Food-related gastrointestinal signals activate caudal brainstem neurons expressing both NMDA and AMPA receptors. *Brain Res.* **915,** 143–154 (2001).
24. Covasa, M., Hung, C. Y., Ritter, R. C. & Burns, G. A. Intracerebroventricular administration of MK-801 increases food intake through mechanisms independent of gastric emptying. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287,** R1462–R1467 (2004).
25. Treece, B. R., Covasa, M., Ritter, R. C. & Burns, G. A. Delay in meal termination follows blockade of *N*-methyl-D-aspartate receptors in the dorsal hindbrain. *Brain Res.* **810,** 34–40 (1998).
26. Pocai, A. *et al.* Restoration of hypothalamic lipid sensing normalizes energy and glucose homeostasis in overfed rats. *J. Clin. Invest.* **116,** 1081–1091 (2006).
27. Caspi, L., Wang, P. Y. & Lam, T. K. A balance of lipid-sensing mechanisms in the brain and liver. *Cell Metab.* **6,** 99–104 (2007).
28. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
29. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
30. Cummings, D. E., Overduin, J. & Foster-Schubert, K. E. Gastric bypass for obesity: mechanisms of weight loss and diabetes resolution. *J. Clin. Endocrinol. Metab.* **89,** 2608–2615 (2004).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank C. Baveghems for technical assistance. This work is supported by a research grant to T.K.T.L. from the Canadian Institute of Health Research (MOP-82701). R.G.-J. is supported by the National Institutes of Health (DK45024). G.J.S. is supported by the National Institutes of Health (DK47208) and the Skirball Institute. T.K.T.L. holds the John Kitson McIvor Endowed Chair in Diabetes Research at the University Health Network and the University of Toronto.

**Author Contributions** P.Y.T.W. conducted and designed experiments, performed data analyses and wrote the manuscript; L.C., C.K.L.L., M.C. and M.A. conducted experiments; X.L. assisted in surgical procedures; P.E.L. and R.G.-J. assisted in LCFA-CoA measurements; G.J.S. assisted in surgical procedures and designed experiments; and T.K.T.L. supervised the project, designed experiments and wrote the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.K.T.L. (tony.lam@uhnres.utoronto.ca).

LETTERS

# A deadenylation negative feedback mechanism governs meiotic metaphase arrest

Eulàlia Belloc[1] & Raúl Méndez[1]

**In vertebrate oocytes, meiotic progression is driven by the sequential translational activation of maternal messenger RNAs stored in the cytoplasm. This activation is mainly induced by the cytoplasmic elongation of their poly(A) tails, which is mediated by the cytoplasmic polyadenylation element (CPE) present in their 3′ untranslated regions[1,2]. In *Xenopus* oocytes, sequential phase-specific translation of CPE-regulated mRNAs is required to activate the maturation-promoting factor, which in turn mediates entry into the two consecutive meiotic metaphases (MI and MII)[3–6]. Here we report a genome-wide functional screening to identify previously unknown mRNAs cytoplasmically polyadenylated at meiotic phase transitions. A significant fraction of transcripts containing, in addition to CPEs, (A + U)-rich element (ARE) sequences (characteristic of mRNAs regulated by deadenylation[7]) were identified. Among these is the mRNA encoding C3H-4, an ARE-binding protein that we find to accumulate in MI and the ablation of which induces meiotic arrest. Our results suggest that C3H-4 recruits the CCR4 deadenylase complex to ARE-containing mRNAs and this, in turn, causes shortening of poly(A) tails. We also show that the opposing activities of the CPEs and the AREs define the precise activation times of the mRNAs encoding the anaphase-promoting complex inhibitors Emi1 and Emi2 during distinct phases of the meiotic cycle. Taken together, our results show that an 'early' wave of cytoplasmic polyadenylation activates a negative feedback loop by activating the synthesis of C3H-4, which in turn would recruit the deadenylase complex to mRNAs containing both CPEs and AREs. This negative feedback loop is required to exit from metaphase into interkinesis and for meiotic progression.**

Cytoplasmic polyadenylation requires two elements in the 3′ untranslated regions (UTRs) of responding mRNAs: the polyadenylation signal (Hex, also known as Hhex) and the nearby CPE, which recruits the CPE-binding protein (CPEB). However, the activation of CPE-containing mRNAs does not occur *en masse* at any one time (see Supplementary Fig. 1). Instead, the polyadenylation of specific mRNAs is temporally regulated[3,4] by two sequential phosphorylations of CPEB. First, phosphorylation of CPEB by aurora A kinase[5] at the prophase I (PI)–MI transition induces a first wave of 'early' or Cdc2-independent cytoplasmic polyadenylation[8], required for the PI–MI transition. Then, CPEB phosphorylation by Cdc2 at MI triggers a second wave of 'late' or Cdc2-dependent polyadenylation[8], required for the MI–MII transition[6,9]. However, a less understood third wave of polyadenylation takes place during interkinesis targeting mRNAs (such as those encoding cyclin E[10] and Wee1, ref. 11) that have 'late' CPEs plus an additional U-rich sequence that has been defined as an embryonic CPE[12]. Although neither the mechanism

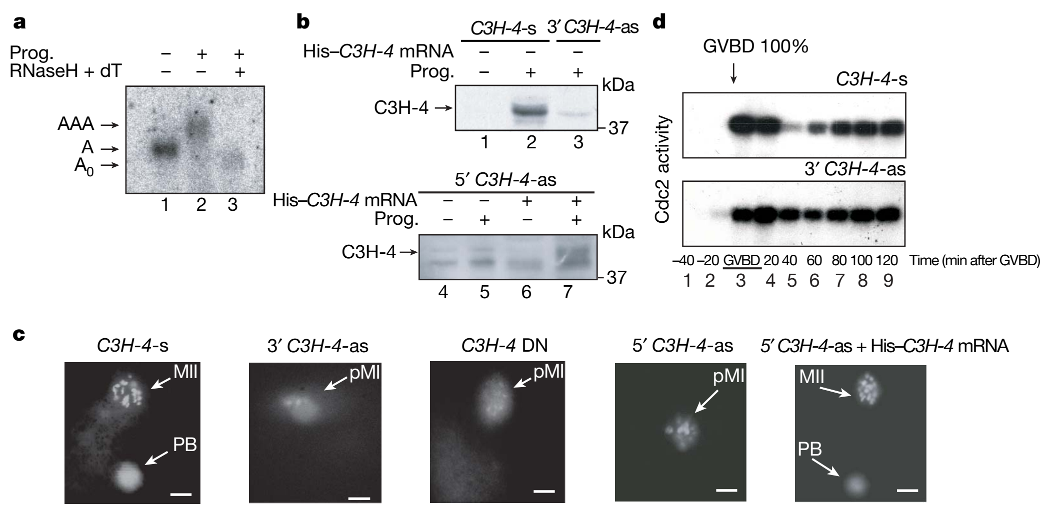



**Figure 1 | Polyadenylation-induced translation of *C3H-4* mRNA is required to exit MI. a**, Northern blot of *C3H-4* mRNA from untreated or progesterone (Prog.)-stimulated oocytes. Where indicated, the RNA was incubated with oligo(dT) and RNaseH. **b**, Oocytes were microinjected with *C3H-4* antisense oligonucleotides (3′ *C3H-4*-as or 5′ *C3H-4*-as), *C3H-4* sense oligonucleotides (*C3H-4*-s) or C3H-4-encoding mRNA (+ His–*C3H-4* mRNA) and were then treated with progesterone as indicated. Expression of C3H-4 was analysed by western blot. **c**, Oocytes were fixed and stained for chromosome visualization 6 h after GVBD. The MII metaphase plate (MII), polar body (PB) and the MI pseudometaphase plate (pMI) are indicated. Scale bars represent 20 μm. DN, dominant negative C3H-4 variant. **d**, The oocytes were collected at the indicated times after progesterone stimulation and analysed for Cdc2 kinase activity.

[1]Centre for Genomic Regulation (CRG), Pompeu Fabra University (UPF), C/Dr Aiguader 88, 08003, Barcelona, Spain.

dictating this third wave of 'late–late' cytoplasmic polyadenylation nor its meiotic function have been defined, it could be required to establish the MII arrest.

To identify the maternal mRNAs controlling meiotic progression and to elucidate how their translational regulation is temporally defined during the meiotic phase transitions, we designed a genome-wide functional screening for mRNAs polyadenylated in the course of the PI–MI and the MI–MII transitions (Supplementary Fig. 2a–c and Supplementary Tables 1 and 2). In addition to the Hex and the CPEs, 57% of these UTRs contained class I AREs and 38% of the UTRs contained U-rich sequences resembling class III AREs or embryonic CPEs[13]. Depletion of five of the newly identified mRNAs confirmed that their translation was required for meiotic progression (Supplementary Fig. 2e). Because more than half of the UTRs identified in our screening contained AREs, we focused our attention in C3H-4, an ARE-binding zinc-finger protein that is a member of the Tristetraprolin family of proteins[14,15]. We first confirmed that the endogenous mRNA encoding C3H-4 was indeed deadenylated in PI-arrested oocytes and was subsequently polyadenylated in response to progesterone (Fig. 1a) in a Mos/Cdc2-independent manner (Supplementary Fig. 3). C3H-4 protein was synthesized in response to progesterone (Fig. 1b and Supplementary Fig. 4). These data indicate that C3H-4 is encoded by a maternal mRNA activated by 'early' cytoplasmic polyadenylation during the PI–MI transition.

We next aimed to determine whether C3H-4 synthesis was required for meiotic progression beyond MI. Depletion of C3H-4 (Fig. 1b) or overexpression of a C3H-4 dominant-negative variant induced partial chromosomal decondensation in the absence of a polar body (Fig. 1c) and abnormal morphology, consistent with a defect in the MI–MII transition (Supplementary Fig. 5). This phenotype was rescued by overexpressing C3H-4 from a microinjected mRNA not targeted by the antisense oligonucleotide (Fig. 1b, c). To better define at which step of meiosis the *C3H-4* antisense microinjected oocytes were arrested, we performed a time course of Cdc2 activity (Fig. 1d). In the control oocytes, Cdc2 activity sharply decreased after germinal vesicle breakdown (GVBD) and increased again at MII, whereas, in the C3H-4-depleted oocytes, Cdc2 activity was maintained at high levels during the whole time course. Oocyte maturation and Cdc2 activation took place about an hour earlier in C3H-4-depleted oocytes than in control oocytes (data not shown). Altogether, these results show that C3H-4 depletion accelerates maturation and induces meiotic arrest in a pseudo-metaphase state[16,17] during the first meiotic division. We conclude that C3H-4 opposes maturation-promoting factor (MPF) activation and is required for exit of the first meiotic metaphase.

We next sought to determine whether C3H-4 could mediate deadenylation of ARE-containing mRNAs. Indeed, a polyadenylated fragment of the tumour necrosis factor α (*TNF-α*) 3′ UTR, which contains well characterized AREs and binds C3H-4 (ref. 15), was rapidly deadenylated on MI entry, concomitant with C3H-4 synthesis (Fig. 2a and Supplementary Fig. 6). Ablation of *C3H-4* mRNA indicated that *TNF-α* deadenylation requires C3H-4 (Fig. 2a). We next aimed to determine whether C3H-4 recruited a deadenylase to the ARE-containing mRNAs. For this, we focused in the CCR4–Not deadenylation complex, which is known to mediate deadenylation of cell-cycle-related ARE-containing mRNAs in vertebrates[18,19]. Indeed, CCR4 co-immunoprecipitated with C3H-4, and vice versa, in extracts from MII oocytes (Fig. 2b). Because C3H-4 was detected as a doublet in MII (Fig. 2b), we tested whether this protein was phosphorylated in the course of meiosis, suggesting also a post-translational regulation of its activity, in agreement with the fact that we did not detect any clear effect of overexpressing C3H-4 (data not shown). Indeed, overexpressed tagged C3H-4 displayed slower gel mobility in MI than in PI and an intermediate mobility in MII (Fig. 2c)—an effect that was reverted by treatment with alkaline phosphatase (Supplementary Fig. 7). When C3H-4 is used as substrate in an *in vitro* phosphorylation assay with extracts from oocytes at different meiotic stages as the source of kinase, we detected phosphorylation of C3H-4 in PI and MII and hyperphosphorylation, shown as a slow-migrating band, in MI (Fig. 2d). These observations, together with the fact that CCR4 seems to co-immunoprecipitate mainly the slow-migrating form of C3H-4, could reflect a post-translational activation of C3H-4 in MI and inactivation for MII entry.

To identify C3H-4-regulated mRNAs, we used a candidate approach. Because C3H-4 mediates deadenylation in MI, and therefore translational inhibition, we searched for ARE- and CPE-containing mRNAs encoding proteins that, when overexpressed in MI, caused a similar meiotic arrest. Two of the best candidates were the anaphase-promoting complex (APC) inhibitors *emi1* and *emi2*. Emi1 is transiently expressed at low levels during the PI–MI transition, and its degradation in MI is required for progression beyond prometaphase-I (refs 20–22). Emi2 is synthesized during interkinesis[23,24] and is required for the establishment and maintenance of the CSF[25–27].

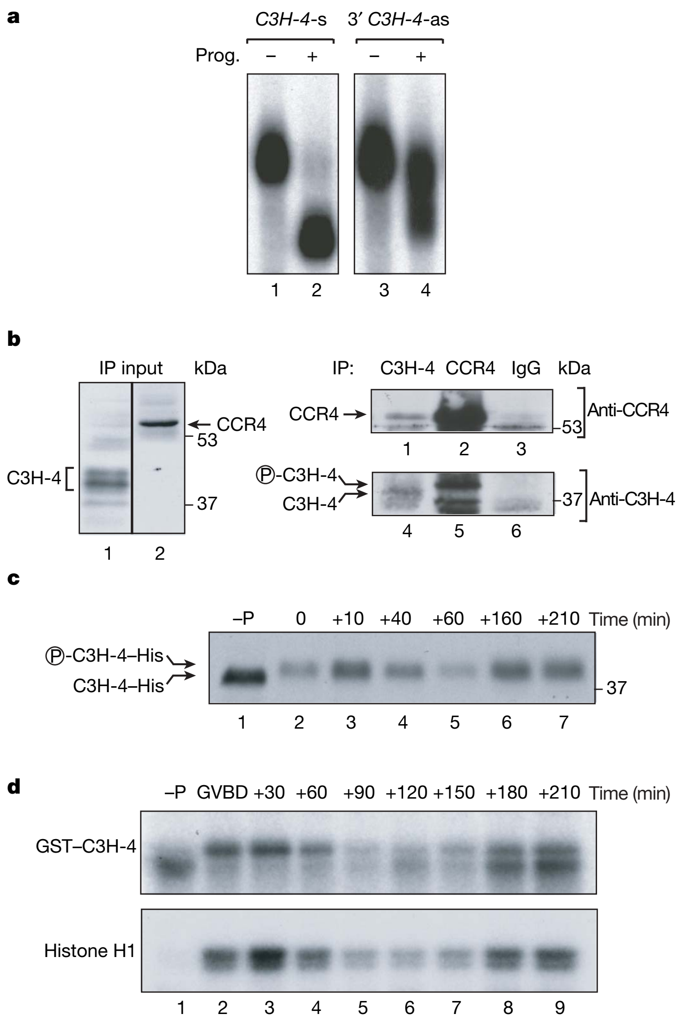



**Figure 2 | C3H-4 mediates ARE-dependent deadenylation in oocytes by recruiting the CCR4 deadenylase to the mRNAs.** **a**, Oocytes injected with either *C3H-4* sense (s) or *C3H-4* antisense (as) oligonucleotides were microinjected with a radiolabelled polyadenylated *TNF-α* RNA probe, incubated in the absence or presence of progesterone (Prog.), and then total RNA was extracted and analysed by gel electrophoresis followed by autoradiography. **b**, Cytoplasmic extracts from mature oocytes, or the corresponding fractions immunoprecipitated (IP) with anti-C3H-4 and anti-CCR4 antibodies or control immunoglobulin (Ig)Gs, were analysed by western blot with either C3H-4 or CCR4 antibodies. **c**, His–*C3H-4*-mRNA-microinjected oocytes not treated with progesterone (−P), or treated and collected at the indicated times after maturation, were analysed by western blot for C3H-4. **d**, *In vitro* kinase assays of oocyte extracts collected at the indicated times and using either GST–C3H-4 or histone H1 as substrates.

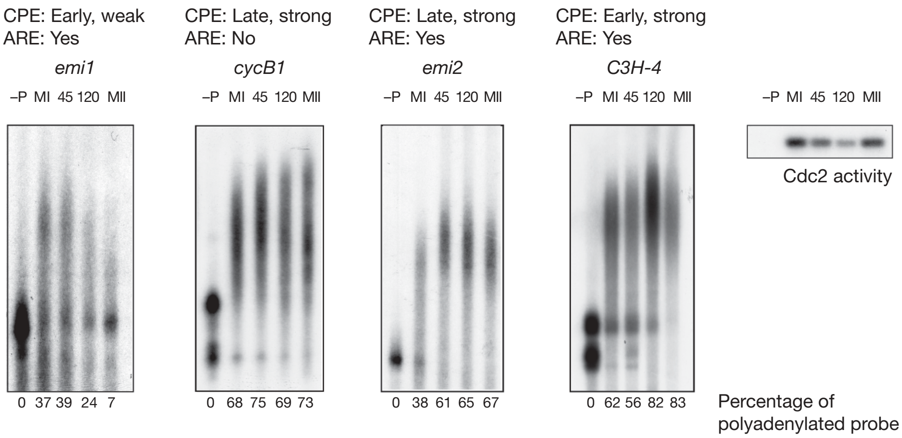



**Figure 3 | The polyadenylation profile is defined by the CPE arrangement and the presence of AREs.** The indicated radiolabelled RNAs were microinjected into oocytes, and then incubated in the presence or absence (−P) of progesterone. At the indicated times after maturation (min), the oocytes were collected and total RNA was extracted and analysed by gel electrophoresis followed by autoradiography. Cdc2 kinase activity at each point was determined as in Fig. 1. The percentage of polyadenylated RNA is indicated.

Premature expression of Emi2 at MI causes MI arrest[23,24]. The *emi1* 3′ UTR contains five potential 'early-weak' CPEs[8] and multiple potential AREs. The *emi2* 3′ UTR contains four 'late-strong' CPEs[8] and multiple potential AREs (Supplementary Fig. 8a). A recent report[24] shows that the 3′ UTR of *emi2* is polyadenylated in a CPE-dependent manner.

To test whether *emi1* and *emi2* mRNAs could be regulated by cytoplasmic changes in their polyadenylation status during meiosis, we compared the polyadenylation kinetics of microinjected labelled probes for the 3′ UTRs of *emi1*, *emi2*, cyclin B1 (which contains strong-late CPEs[8] but no AREs) and *C3H-4* (which contains strong-early CPEs[8] and AREs; Supplementary Fig. 8a). The *emi1* 3′ UTR was weakly polyadenylated during the PI–MI transition and was sharply deadenylated between MI and MII. Cyclin B1 3′ UTR was already maximally polyadenylated at MI, whereas *emi2*, which contains a similar arrangement of CPEs, was not maximally polyadenylated until interkinesis. *C3H-4* was partially polyadenylated in MI and fully polyadenylated in MII (Fig. 3). These data suggest that the specific arrangement of CPEs combined with the presence or absence of AREs defines specific patterns of poly(A) tail length during meiosis.

To determine whether the *emi1* and *emi2* mRNA-specific patterns of polyadenylation were indeed defined by CPEB and C3H-4, we first tested whether U-labelled probes from *emi1* and *emi2* 3′ UTRs were able to bind these proteins. As controls, we used probes derived from *TNF-α* 3′ UTR (which contains AREs but no CPEs) or cyclin B1 3′ UTR (Supplementary Fig. 8a). Ultraviolet crosslinking of cytoplasmic extracts from C3H-4-overexpressing oocytes, followed by C3H-4 immunoprecipitation or CPEB purification, showed that *emi1*, *emi2* and *TNF-α*, but not cyclin B1, probes bound C3H-4 (Fig. 4a). In

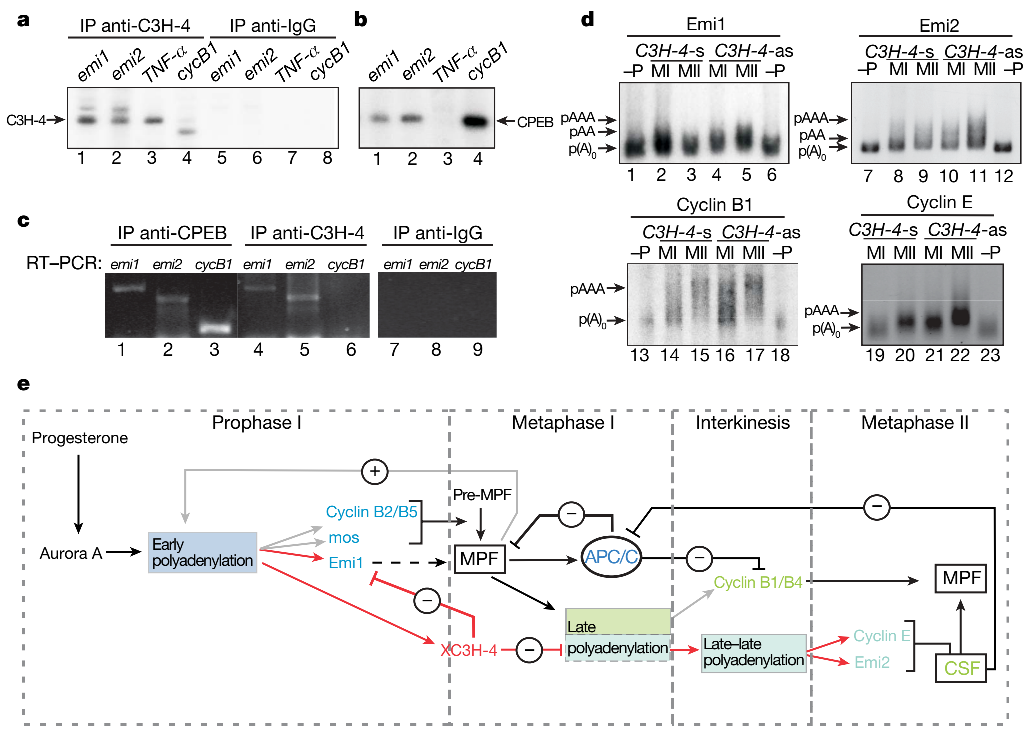



**Figure 4 | The polyadenylation of *emi1* and *emi2*, CPE- and ARE-containing mRNAs, is regulated by CPEB and C3H-4. a**, The indicated radiolabelled RNAs were incubated in oocyte extracts overexpressing C3H-4, were ultraviolet-crosslinked and were then immunoprecipitated with anti-C3H-4 antibody or control IgGs. The immunoprecipitates were analysed by SDS–PAGE followed by autoradiography. **b**, The same RNA probes and ultraviolet-crosslinked extracts were subjected to sepharose chromatography and analysed as in **a**. **c**, Cytoplasmic extracts from MI oocytes were immunoprecipitated with anti-CPEB, anti-C3H-4 and control-IgG antibodies. The immunoprecipitates were analysed for the indicated mRNAs by RT–PCR. **d**, Oocytes microinjected with 3′ *C3H-4* antisense oligonucleotides (*C3H-4*-as) or *C3H-4* sense oligonucleotides (*C3H-4*-s) were treated with progesterone and collected at metaphase I (MI) or metaphase II (MII), or were left untreated (−P). Total RNA was extracted and the polyadenylation status of the indicated mRNAs measured by RNA-ligation-coupled RT–PCR. **e**, Schematic diagram showing the sequential waves of polyadenylation and deadenylation driving meiotic progression.

contrast, CPEB was crosslinked to *emi1*, *emi2* and cyclin B1, but not to *TNF*-$\alpha$, probes (Fig. 4b).

We next sought to determine whether the endogenous mRNAs were also associated to C3H-4 and CPEB by immunoprecipitation with either C3H-4 or CPEB antibodies followed by polymerase chain reaction with reverse transcriptase (RT–PCR) for each mRNA. Indeed, the endogenous *emi1*, *emi2* and cyclin B1 mRNAs were associated with CPEB in MI oocytes (Fig. 4c), whereas *emi1* and *emi2*, but not cyclin B1, mRNAs were bound to C3H-4 (Fig. 4c). Thus, we conclude that *emi1* and *emi2* mRNAs recruit CPEB and C3H-4. To determine whether poly(A) tail length of *emi1* and *emi2* mRNAs was controlled by C3H-4, we analysed their polyadenylation in either control oocytes or C3H-4-depleted oocytes (Fig. 4d). In control oocytes, *emi1* mRNA was deadenylated in PI, polyadenylated in MI and deadenylated again in MII, whereas, in C3H-4-depleted oocytes, *emi1* was not deadenylated in MII but was further polyadenylated. In the same control oocytes, *emi2* mRNA was also deadenylated in PI but maximum polyadenylation was not detected until MII, whereas, in the absence of C3H-4, the mRNA was already polyadenylated in MI. Cyclin E mRNA, another component of the CSF[10], which has a similar arrangement of CPEs and AREs to *emi2*, displayed the same polyadenylation pattern as that of *emi2* mRNA in the presence and absence of C3H-4. As expected, the polyadenylation of cyclin B1 mRNA was not affected by the depletion of C3H-4. Thus, for *emi1*, the weak 'early' polyadenylation at the PI–MI transition was reverted once C3H-4 is synthesized in MI, whereas, for *emi2* and cyclin E, the strong 'late' polyadenylation at MI was displaced to interkinesis by C3H-4, probably as the result of the opposing activities of CPEB and C3H-4 in MI. Accordingly, Emi2 protein synthesis in oocytes expressing C3H-4 takes place during interkinesis/MII, whereas, in the absence of C3H-4, Emi2 is already expressed at high levels in MI (Supplementary Fig. 8b), explaining therefore the MI arrest caused by preventing *C3H-4* mRNA translation.

Taken together, our results show that the 'early' wave of cytoplasmic polyadenylation activates a negative feedback loop (which opposes CPEB activity on mRNAs containing both CPEs and AREs) by activating the synthesis of C3H-4. The effect of the C3H-4-mediated deadenylation on the target mRNAs is in turn modulated by the different arrangements of CPEs. Thus, for an mRNA polyadenylated by the 'early' activation of a weak CPE, C3H-4 inactivates the mRNA after MI. However, for mRNAs containing a late-strong CPE arrangement, C3H-4 causes a delay in the poly(A) tail elongation, generating a third wave of polyadenylation in interkinesis. These sequential waves of polyadenylation and deadenylation drive meiotic progression (Fig. 4e). Thus, 'early' polyadenylation of *mos*, cyclins B2 and B5 and *emi1* would be required to activate and sustain MPF activity in MI, whereas 'late' polyadenylation of cyclins B1 and B4 would be required to sustain an intermediate MPF activity during interkinesis and for the reactivation of MPF in MII. In addition, activation of *emi2* and cyclin E during interkinesis is required to generate CSF. Conversely, 'early' polyadenylation-driven translation of *C3H-4* generates a deadenylation wave that inactivates *emi1* translation in MI allowing the transition to interkinesis, and delays Emi2 and cyclin E synthesis to late interkinesis preventing premature activation of CSF. Sequential waves of polyadenylation and deadenylation ensure irreversible, self-sustained, meiotic phase transitions by controlling discrete states of MPF, APC and CSF activities. The bistability of this process is sustained by multiple translation-dependent positive and negative feedback loops that help to keep the oocyte from slipping rapidly back and forth between cell cycle phases[28–30].

## METHODS SUMMARY

**Oocyte manipulation and microinjection.** Stage VI oocytes were obtained from the ovaries of adult *Xenopus* females as described previously[3]. Maturation was induced by incubating oocytes in Barth's medium containing 10 μM progesterone. DNA staining to assess meiotic progression was performed as described[31], and the Cdc2 kinase activity was assayed as described[6].

**Antisense oligonucleotide and rescue experiment.** To ablate the expression of *C3H-4* mRNA, 18-mer oligonucleotides targeting either the 5′ UTR or the 3′ UTR were designed; the complementary sequence of one of them was used as a control. In each oocyte, 98 ng oligonucleotides was injected. After overnight (16 h) incubation at 18 °C, progesterone was added as described. For the rescue experiment, the 5′ *C3H-4* antisense injected oocytes were used followed by injection of 0.03 pmol of *in vitro* transcribed RNA coding for the ORF and the 3′ UTR of *C3H-4* mRNA. See Methods for further details of the oligonucleotide sequences.

**RNA-ligation-coupled RT–PCR.** This technique was performed as described[32] with minor modifications. See Methods for detailed modifications.

**Antibodies and immunoprecipitations.** The CPEB antibody was a gift from J. D. Richter, the dCCR4 antibody was a gift from E. Wahle and the Emi2 antibody was a gift from T. U. Mayer. The ultraviolet crosslinking followed by immunoprecipitations were performed as described[33], using 50 fmol of *emi1* and *emi2* labelled RNAs and 100 fmol of *TNF-α* and cyclin B1 labelled RNAs. Immunoprecipitations followed by RT–PCR were performed as described[34] (using fresh MI oocyte lysates, 50 oocytes per condition) with the C3H-4 and CPEB antibodies. The protein-bound RNAs were purified by proteinase K digestion followed by phenol–chloroform extraction.

**Polyadenylation and deadenylation assays.** These assays were carried out as described in ref. 8.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 18 December 2007; accepted 6 February 2008.**
**Published online 2 April 2008.**

1. Mendez, R. & Richter, J. D. Translational control by CPEB: a means to the end. *Nature Rev. Mol. Cell Biol.* **2**, 521–529 (2001).
2. Richter, J. D. CPEB: a life in translation. *Trends Biochem. Sci.* **32**, 279–285 (2007).
3. de Moor, C. H. & Richter, J. D. The Mos pathway regulates cytoplasmic polyadenylation in *Xenopus* oocytes. *Mol. Cell. Biol.* **17**, 6419–6426 (1997).
4. Ballantyne, S., Daniel, D. L. Jr & Wickens, M. A dependent pathway of cytoplasmic polyadenylation reactions linked to cell cycle control by c-mos and CDK1 activation. *Mol. Biol. Cell* **8**, 1633–1648 (1997).
5. Mendez, R. *et al.* Phosphorylation of CPE binding factor by Eg2 regulates translation of *c-mos* mRNA. *Nature* **404**, 302–307 (2000).
6. Mendez, R., Barnard, D. & Richter, J. D. Differential mRNA translation and meiotic progression require Cdc2-mediated CPEB destruction. *EMBO J.* **21**, 1833–1844 (2002).
7. Voeltz, G. K. & Steitz, J. A. AUUUA sequences direct mRNA deadenylation uncoupled from decay during *Xenopus* early development. *Mol. Cell. Biol.* **18**, 7537–7545 (1998).
8. Piqué, M., López, J. M., Foissac, S., Guigó, R. & Méndez, R. A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434–448 (2008).
9. Mendez, R., Murthy, K. G., Ryan, K., Manley, J. L. & Richter, J. D. Phosphorylation of CPEB by Eg2 mediates the recruitment of CPSF into an active cytoplasmic polyadenylation complex. *Mol. Cell* **6**, 1253–1259 (2000).
10. Rempel, R. E., Sleight, S. B. & Maller, J. L. Maternal *Xenopus* Cdk2–cyclin E complexes function during meiotic and early embryonic cell cycles that lack a G1 phase. *J. Biol. Chem.* **270**, 6843–6855 (1995).
11. Charlesworth, A., Welk, J. & MacNicol, A. M. The temporal control of *Wee1* mRNA translation during *Xenopus* oocyte maturation is regulated by cytoplasmic polyadenylation elements within the 3′-untranslated region. *Dev. Biol.* **227**, 706–719 (2000).
12. Simon, R. & Richter, J. D. Further analysis of cytoplasmic polyadenylation in *Xenopus* embryos and identification of embryonic cytoplasmic polyadenylation element-binding proteins. *Mol. Cell. Biol.* **14**, 7867–7875 (1994).
13. Wu, L., Good, P. J. & Richter, J. D. The 36-kilodalton embryonic-type cytoplasmic polyadenylation element-binding protein in *Xenopus laevis* is ElrA, a member of the ELAV family of RNA-binding proteins. *Mol. Cell. Biol.* **17**, 6402–6409 (1997).
14. De, J. *et al.* Identification of four CCCH zinc finger proteins in *Xenopus*, including a novel vertebrate protein with four zinc fingers and severely restricted expression. *Gene* **228**, 133–145 (1999).
15. Lai, W. S., Carballo, E., Thorn, J. M., Kennington, E. A. & Blackshear, P. J. Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to AU-rich elements and destabilization of mRNA. *J. Biol. Chem.* **275**, 17827–17837 (2000).
16. Parry, D. H., Hickson, G. R. & O'Farrell, P. H. Cyclin B destruction triggers changes in kinetochore behavior essential for successful anaphase. *Curr. Biol.* **13**, 647–653 (2003).
17. Yang, Z. *et al.* Silencing mitosin induces misaligned chromosomes, premature chromosome decondensation before anaphase onset, and mitotic cell death. *Mol. Cell. Biol.* **25**, 4062–4074 (2005).
18. Collart, M. A. Global control of gene expression in yeast by the Ccr4–Not complex. *Gene* **313**, 1–16 (2003).

19. Morita, M. *et al.* Depletion of mammalian CCR4b deadenylase triggers increment of the p27Kip1 mRNA level and impairs cell growth. *Mol. Cell. Biol.* 13, 4980–4990 (2007).
20. Reimann, J. D. *et al.* Emi1 is a mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex. *Cell* 105, 645–655 (2001).
21. Margottin-Goguet, F. *et al.* Prophase destruction of Emi1 by the SCF(βTrCP/Slimb) ubiquitin ligase activates the anaphase promoting complex to allow progression beyond prometaphase. *Dev. Cell* 4, 813–826 (2003).
22. Tung, J. J. *et al.* A role for the anaphase-promoting complex inhibitor Emi2/XErp1, a homolog of early mitotic inhibitor 1, in cytostatic factor arrest of *Xenopus* eggs. *Proc. Natl Acad. Sci. USA* 102, 4318–4323 (2005).
23. Ohe, M., Inoue, D., Kanemori, Y. & Sagata, N. Erp1/Emi2 is essential for the meiosis I to meiosis II transition in *Xenopus* oocytes. *Dev. Biol.* 303, 157–164 (2007).
24. Tung, J. J., Padmanabhan, K., Hansen, D. V., Richter, J. D. & Jackson, P. K. Translational unmasking of Emi2 directs cytostatic factor arrest in meiosis II. *Cell Cycle* 6, 725–731 (2007).
25. Inoue, D., Ohe, M., Kanemori, Y., Nobui, T. & Sagata, N. A direct link of the Mos–MAPK pathway to Erp1/Emi2 in meiotic arrest of *Xenopus laevis* eggs. *Nature* 446, 1100–1104 (2007).
26. Nishiyama, T., Ohsumi, K. & Kishimoto, T. Phosphorylation of Erp1 by p90rsk is required for cytostatic factor arrest in *Xenopus laevis* eggs. *Nature* 446, 1096–1099 (2007).
27. Liu, J., Grimison, B. & Maller, J. L. New insight into metaphase arrest by cytostatic factor: from establishment to release. *Oncogene* 26, 1286–1289 (2007).
28. Ferrell, J. E. Jr. Building a cellular switch: more lessons from a good egg. *Bioessays* 21, 866–870 (1999).
29. Ferrell, J. E. Jr. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* 14, 140–148 (2002).
30. Brandman, O., Ferrell, J. E. Jr, Li, R. & Meyer, T. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* 310, 496–498 (2005).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank J. D. Richter for the anti-CPEB antibody; E. Wahle for the dCCR4 antibody; T. U. Mayer for the Emi2 antibody; and M. Fernández, members of the Méndez laboratory, J. Valcarcel and other colleagues from the Gene Expression Program for advice and for critically reading the manuscript. This work was supported by grants from the MEC, Fundación 'La Caixa' and Fundació 'Marató de TV3'. R.M. is a recipient of a contract from the 'Programa Ramon y Cajal' (MEC).

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.M. (raul.mendez@crg.es).

## METHODS

**Purification of poly(A)(+80) RNA with poly(U)-sepharose.** A *Xenopus* oligo(dT)-primed complementary DNA oocyte library was *in vitro* transcribed with the T3 mMessage mMachine kit (Ambion), and then the RNAs were treated with RNase H and oligo(dT) to remove their poly(A) tail. These RNAs were microinjected (46 nl per oocyte) into stage VI oocytes that were treated with progesterone, and, when mature, total RNA was extracted with Ultraspec (Biotecx) and the poly(A)(+80) RNA fraction was purified by poly(U) chromatography. Total RNA was incubated with sample buffer (0.01 M Tris-HCl, pH 7.5, 1 mM EDTA, 1% SDS) for 5 min at 65 °C, the sample was chilled on ice and binding buffer was added (0.05 M Tris-HCl, pH 7.5, 0.7 M NaCl, 10 mM EDTA, 25% (v/v) formamide), and then the sample was loaded into the poly(U)-agarose (Sigma) chromatography column and incubated for 30 min at room temperature (22–23 °C) with agitation; the column had been previously washed with swelling buffer (0.05 M Tris-HCl, pH 7.5, 1 M NaCl) and packed. Next, the column containing the sample was washed three times at 25 °C and three times at 55 °C with washing buffer (0.05 M Tris-HCl, pH 7.5, 0.1 M NaCl, 10 mM EDTA, 25% formamide) and the remaining poly(A)(+80) RNA was eluted with elution buffer (0.05 M HEPES, pH 7, 10 mM EDTA, 90% (v/v) formamide). The eluted RNA was precipitated after phenol–chloroform extraction. Reverse transcription (RevertAid M-MuLV retrotranscriptase, Fermentas) of the purified RNA was performed with 3′ Race oligonucleotides (TAATACGACTCACTATAGGGCGGATCCTTTTTTTTTTTTTTTTTTTTTTTTTVN), and then PCR (Ecotaq polymerase, Ecogen) with T7-antisense (GTAATACGACTCACTATAGGGC) and T3zorn (GGAAATTAACCCTCACTAAAGGGCAGAATAAACGCTCAACTTTGGC) primers was performed to amplify the library. To clone the products into a pBSK vector, the PCR products were digested with BamHI and EcoRI.

We sequenced the 253 clones from the early library and 246 of the 2,620 clones from the late library. After correcting for redundancy, we found 56 UTRs from 'early' polyadenylated mRNAs and 55 from mRNAs polyadenylated in MI.

**Antisense oligonucleotides for meiotic progression screening.** The following antisense oligonucleotides were used: *hsp90*-antisense (GGTGCAGAGAAAGGATAC); *pp2c*-antisense (CCAATCCCCCCTCCAATG); *X71067*-antisense (AATTTTCGATAGTCCCTC); and *tyrP4a2*-antisense (GGTGGGGGCAATGCTTAG).

**RNA-ligation-coupled RT–PCR.** Total oocyte RNA (4 μg) from pools of six oocytes was ligated to 0.4 μg SP2 anchor primer (5′-P-GGTCACCTCTGATCTGGAAGCGAC-$NH_2$-3′) in a 10 μl reaction using T4 RNA ligase (New England Biolabs) according to the manufacturer's directions. The whole 10 μl RNA ligation reaction was used in a 50 μl reverse transcription, using the RevertAid M-MuLV retrotranscriptase (Fermentas) according to manufacturer's directions with 0.4 μg ASP2T (GTCGCTTCCAGATCAGAGGTGACCTTTTT) as the reverse primer. One microlitre of this cDNA preparation was used in each 25 μl PCR reaction using EcoTaq polymerase (Ecogen) and [$^{32}$P] dATP to label the product. The specific primers used were *emi1*-sense (GCAAACAGAATTTACGGAGGTTATAG), *emi2*-sense (GGGGCTTTGGGGGTGTGTATGATCCG), 3′ *C3H-4*-sense (GCCAGCAGTTGAGAAGCC), cyclin B1 (GGAGATCTTGTTGGCACCATGTGCTTC) and cyclin E (GGAGGATGCTACTTGATGGAGAT). The PCR products were resolved in polyacrylamide gels and autoexposed.

**Cloning and mutagenesis.** *C3H-4* (Gene Bank accession number AF061983) cDNA was cloned by RT–PCR (using the RevertAid M-MuLV Reverse Transcriptase from Fermentas for RT, and Taq plus precision polymerase from Stratagene for PCR) with *C3H-4*-sense (GGGAATTGCGAGATATCAAATGACAGTCTGG) and *C3H-4*-antisense (GGGTCGACGTATAAAGTTAGCCGAGGACAG) primers, from total *Xenopus* oocyte RNA. The PCR product was cloned into a pGEX4T-3 expression vector, and into pBSK + pA73 vector for *in vitro* transcription.

The *TNF-α* construct was generated from the IMAGE clone 718000 GenBank accession number AA261175 by PCR with the *TNF*-sense (GGAGATCTCAATGCACAGCTTTCCTCAC) and the *TNF*-antisense (CCGGATCCTTCCAACACTGGGTCTTCAG) oligonucleotides. The PCR product was cloned into a plasmid after the T3 promoter for transcription and before a poly(A) tail of 73 nucleotides to obtain a *TNF*-αp$(A)_{73}$ RNA.

The *emi1* (AF319594) and *emi2* (AF450296) 3′ UTRs were cloned into pBSK by performing a PCR from the IMAGE clone 6641423 GenBank accession number BU914851 and IMAGE clone 3405254 GenBank accession number BG486873, respectively, with specific primers containing the SmaI restriction site in the sense primer and the EcoRI restriction site in the antisense primer. The plasmid and the PCR products were digested with SmaI and EcoRI and were ligated.

**Overexpressed C3H-4 *Xenopus* oocytes extracts.** Five-hundred stage VI oocytes were injected with 100 ng of *in vitro* transcribed RNA with the coding sequence of C3H-4 and a poly(A) tail of 73 nucleotides. After overnight incubation at 18 °C, the oocytes were mixed with other 3,000 non-injected oocytes, and then the oocytes were rinsed twice with extract buffer (10 mM HEPES-KOH, pH 7,5, 100 mM KCl, 1 mM $MgCl_2 \cdot 6H_2O$, 50 mM sucrose supplemented with Complete-EDTA free protease inhibitors from Roche). The oocytes were packed at 2,000*g* for 5 min, the excess of buffer was removed, and the sample was centrifuged at 13,562*g* with maximum acceleration and no-brake deceleration in a Beckman Coulter OptiMax Ultracentrifuge, from which the cytoplasmic fraction was recovered.

**Antisense oligonucleotide and rescue experiments.** To ablate the expression of specific mRNAs, we designed 18-mer oligonucleotides targeting either the 5′ UTR or the 3′ UTR of *C3H-4* mRNA: 5′ *C3H-4*-antisense (5′-CCAGCTTTGACTCCCCAT-3′), 3′ *C3H-4*-antisense (5′-GGCTTCTCAACTGCTGGC-3′) and the complementary sequence as a control 3′ *C3H-4*-sense (5′-GCCAGCAGTTGAGAAGCC-3′). In each oocyte, 98 ng of oligonucleotides was injected. After overnight incubation at 18 °C, progesterone was added as described. For the rescue experiment, after overnight incubation of 5′ *C3H-4* antisense-injected oocytes, 0.03 pmol of *in vitro* transcribed RNA coding for the ORF and the 3′ UTR of *C3H-4* mRNA were injected into the oocytes, and 1 h later progesterone was added.

**Antibodies and immunoprecipitations.** C3H-4 antibodies were raised in rabbits against the C3H-4 124_137 peptide (CSLPPRRYGGPYRER) and were affinity-purified with agarose beads (MicroLink Protein Coupling kit, Pierce) coupled to the immunogenic peptide, as described by the manufacturer.

Ultraviolet crosslinking followed by immunoprecipitations were performed as described[33], using 50 fmol of *emi1* and *emi2* labelled RNAs and 100 fmol of *TNF-α* and cyclin B1 labelled RNAs. For the purification of CPEB in Fig. 4b, we took advantage of its ability to bind sepharose in low-stringency conditions; ultraviolet-crosslinked extracts were incubated with sepharose beads in NET-N buffer (20 mM Tris-HCl, pH 8, 1 mM EDTA, 100 mM NaCl, 0.5% NP-40) for 2 h at 4 °C and then washed with NET-N 175 mM NaCl and eluted with sample buffer (200 mM Tris-HCl, pH 6.8, 40% glycerol, 8% SDS, 20 mM DTT). Immunoprecipitations followed by RT–PCR were performed as described[34] with fresh MI oocyte lysates (50 oocytes per condition), C3H-4 affinity-purified antibody and CPEB antibody. The protein-bound RNAs were purified by proteinase K digestion followed by phenol–chloroform extraction. Half of the total RNA extracted was used for the retrotranscription, performed with the 3′ Race primer (TAATACGACTCACTATAGGGCGGATCCTTTTTTTTTTTTTTTTTTTTTTTTTVN) with the mMLuV reverse transcriptase from Fermentas following the manufacturer's instructions. A twentieth part of the cDNA was used for each specific PCR (EcoTaq polimerase, Ecogen) with a common T7-antisense primer (GTAATACGACTCACTATAGGGC) and with the following specific primers; *emi1*-sense (GCAAACAGAATTTACGGAGGTTATAG), *emi2*-sense (GGGGCTTTGGGGGTGTGTATGATCCG) and cyclin B1 (GGAGATCTTGTTGGCACCATGTGCTTC).

31. Castro, A., Mandart, E., Lorca, T. & Galas, S. Involvement of Aurora A kinase during meiosis I–II transition in *Xenopus* oocytes. *J. Biol. Chem.* **278,** 2236–2241 (2003).
32. Charlesworth, A., Cox, L. L. & MacNicol, A. M. Cytoplasmic polyadenylation element (CPE)- and CPE-binding protein (CPEB)-independent mechanisms regulate early class maternal mRNA translational activation in *Xenopus* oocytes. *J. Biol. Chem.* **279,** 17650–17659 (2004).
33. Hake, L. E. & Richter, J. D. CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. *Cell* **79,** 617–627 (1994).
34. Aoki, K., Matsumoto, K. & Tsujimoto, M. *Xenopus* cold-inducible RNA-binding protein 2 interacts with ElrA, the *Xenopus* homolog of HuR, and inhibits deadenylation of specific mRNAs. *J. Biol. Chem.* **278,** 48491–48497 (2003).

# LETTERS

## Crystal structure of the λ repressor and a model for pairwise cooperative operator binding

Steven Stayrook[1], Peera Jaru-Ampornpan[1]†, Jenny Ni[1], Ann Hochschild[2] & Mitchell Lewis[1]

**Bacteriophage λ has for many years been a model system for understanding mechanisms of gene regulation[1]. A 'genetic switch' enables the phage to transition from lysogenic growth to lytic development when triggered by specific environmental conditions. The key component of the switch is the *c*I repressor, which binds to two sets of three operator sites on the λ chromosome that are separated by about 2,400 base pairs (bp)[2,3]. A hallmark of the λ system is the pairwise cooperativity of repressor binding[4]. In the absence of detailed structural information, it has been difficult to understand fully how repressor molecules establish the cooperativity complex. Here we present the X-ray crystal structure of the intact λ *c*I repressor dimer bound to a DNA operator site. The structure of the repressor, determined by multiple isomorphous replacement methods, reveals an unusual overall architecture that allows it to adopt a conformation that appears to facilitate pairwise cooperative binding to adjacent operator sites.**

The binding of λ repressor to operator regions $O_R$ and $O_L$ on the phage chromosome is required during lysogenic growth, when the integrated prophage replicates as part of the host chromosome. This binding serves two vital functions: it prevents transcription of the viral genes necessary for lytic growth and stimulates transcription of the *c*I gene (encoding the repressor). To simultaneously repress and activate transcription, a pair of repressor dimers binds cooperatively to adjacent operator sites in $O_R$, thereby preventing transcription from lytic promoter $P_R$ and stimulating transcription from lysogenic promoter $P_{RM}$. Similarly, cooperative binding to adjacent operator sites in $O_L$ prevents transcription from lytic promoter $P_L$. Finally, long-range interactions between the cooperatively bound repressor dimers at $O_R$ and $O_L$ directly link the regulatory events occurring at these two loci[5].

Although detailed structural information is available for the isolated domains of the *c*I repressor[6,7], the intact molecule has not been structurally characterized. To understand more completely how this repressor functions and to better appreciate the structural basis of cooperativity, we have determined the structure of an intact dimeric repressor bound to a specific 17-bp operator sequence. The repressor–operator complex was crystallized using a repressor molecule that bears the single amino-acid substitution D197G. This mutant repressor forms a functional dimer that binds to single operator sites with the same affinity as the wild-type repressor but is incapable of participating in the dimer–dimer interactions that mediate cooperative DNA binding[8]. The crystal structure was determined by multiple isomorphous replacement methods; although the crystals diffract only to a modest resolution (3.9 Å), the connectivity was readily apparent (Supplementary Fig. 1). The excellent quality of the experimentally phased maps allowed us to position the amino (N)-terminal domains (NTDs) and carboxy (C)-terminal domains (CTDs) with extremely high certainty and to establish the connectivity of the polypeptide chain between these two domains. Because of the limited resolution of the diffraction data and the absence of strong side-chain information, the model of the intact

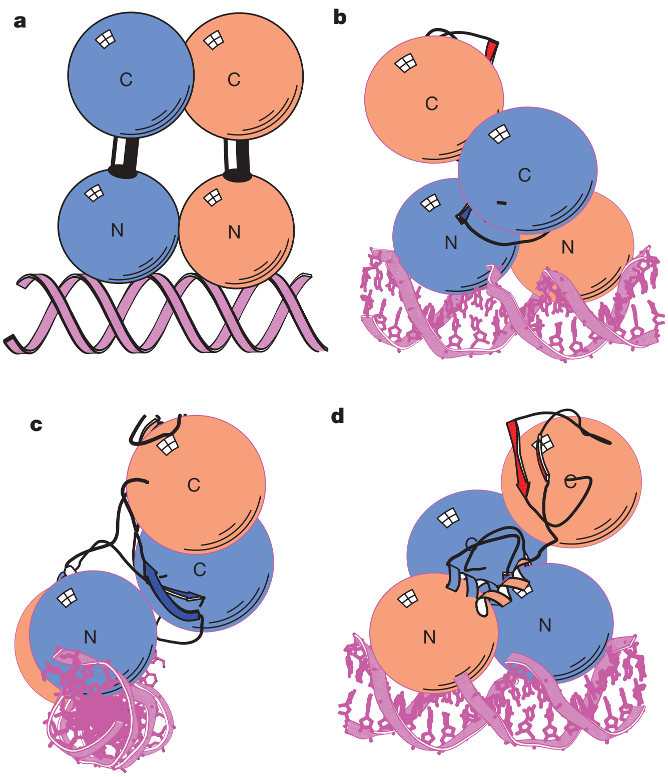



**Figure 1 | Schematic representations of the dimeric repressor bound to the operator.** One subunit is coloured salmon (CSR, red), the other subunit is blue (CSR, dark blue). **a**, Cartoon representation of the repressor bound to the operator. The NTD (N) interacts with the operator DNA, and the CTD (C) self-associates and forms the dimer contacts. The two domains are connected by a linker that tethers the NTD to the CTD. **b**, Relative orientation of the NTDs and CTDs as observed in the crystal structure, as well as the positions of the CSRs. The radii of the spheres were determined based on the actual size of the domains; the centre of each sphere corresponds to the domain's centre of mass. The blue NTD and CTD associate through the CSR. In contrast, the salmon NTD and CTD do not interact with one another. **c**, The complex is rotated by 90° and viewed down the axis of the DNA. In this orientation it is clear that the CSR stabilizes the interactions between the blue NTD and CTD (see also Fig. 2c). **d**, The complex is rotated by 90° with respect to **c** or 180° with respect to **b**. This orientation provides a view of the CSR for the salmon-coloured subunit and illustrates that the NTD and the CTD of this molecule are not in contact. Also shown are the C-terminal helices from the NTDs, to demonstrate the connectivity.

[1]Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, 37th and Hamilton Walk, Philadelphia, Pennsylvania 19102-6059, USA. [2]Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. †Present address: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA.

repressor was refined using tight structural restraints (see Supplementary Data).

The structure of λ repressor is conceptually consistent with the textbook cartoons that depict it as a dumbbell[1] (Fig. 1a). The full-length repressor is composed of two structurally distinct domains that are tethered by a protease sensitive connector[9]. The NTD (residues 1–92), which contains a helix–turn–helix DNA-binding motif, is a compact α-helical domain that weakly self-associates to form a dimer. Dimers of the NTD recognize and bind to the operator sequences using this helix–turn–helix motif[6,10]. The CTD (residues 132–236) is a highly twisted β-sheet structure that is responsible for establishing the essential dimer contacts and for mediating the higher-order dimer–dimer interactions that underlie cooperative binding to the DNA[7]. In addition, the CTD performs a self-cleavage reaction, which is triggered when the lysogenic cell suffers DNA damage and depends upon an activated form of the bacterial RecA protein[11,12]. This self-cleavage reaction inactivates the repressor by

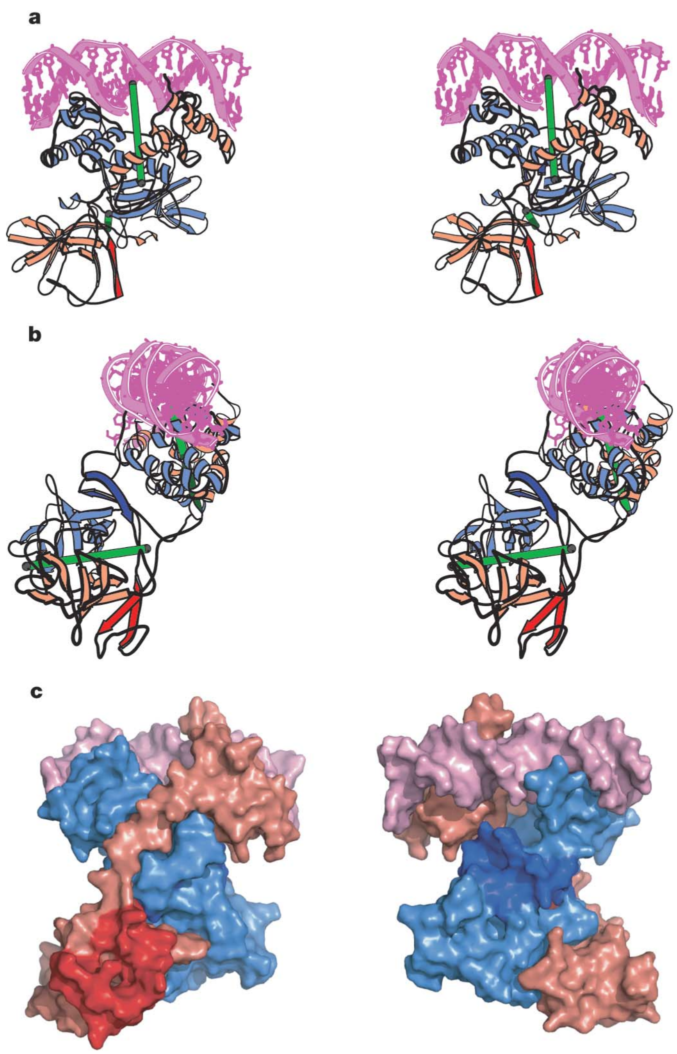



**Figure 2 | Stereo and space-filling views of the dimeric repressor bound to an operator fragment.** Although the NTDs and the CTDs are related by two-fold axes of symmetry, the intact repressor does not have a local dyad. The symmetry axes of the individual domains are shown as green rods (**a**, **b**). The CSRs are coloured dark blue (blue subunit) and red (salmon subunit). **a**, The repressor is oriented such that the principal axis of the DNA is in the plane of the page. The two-fold axis of the NTDs is essentially in the plane of the page, whereas the dyad axis of the CTDs is perpendicular to the page. The blue repressor subunit is more compact and the salmon subunit is elongated. **b**, The complex is rotated 90° and viewed down the major axis of the DNA. This orientation illustrates the asymmetry and that the blue CSR bridges the NTD and the CTD of this subunit. The two-fold axis of the CTDs also relates the red and blue CSRs. **c**, Space-filling representations of the dimeric repressor related by a 180° rotation. Views illustrate that the CSR of the blue molecule (dark blue) creates a bridge between the NTD and the CTD (right image) and that the CSR of the salmon-coloured molecule (red) interacts only with the salmon-coloured CTD (left image).

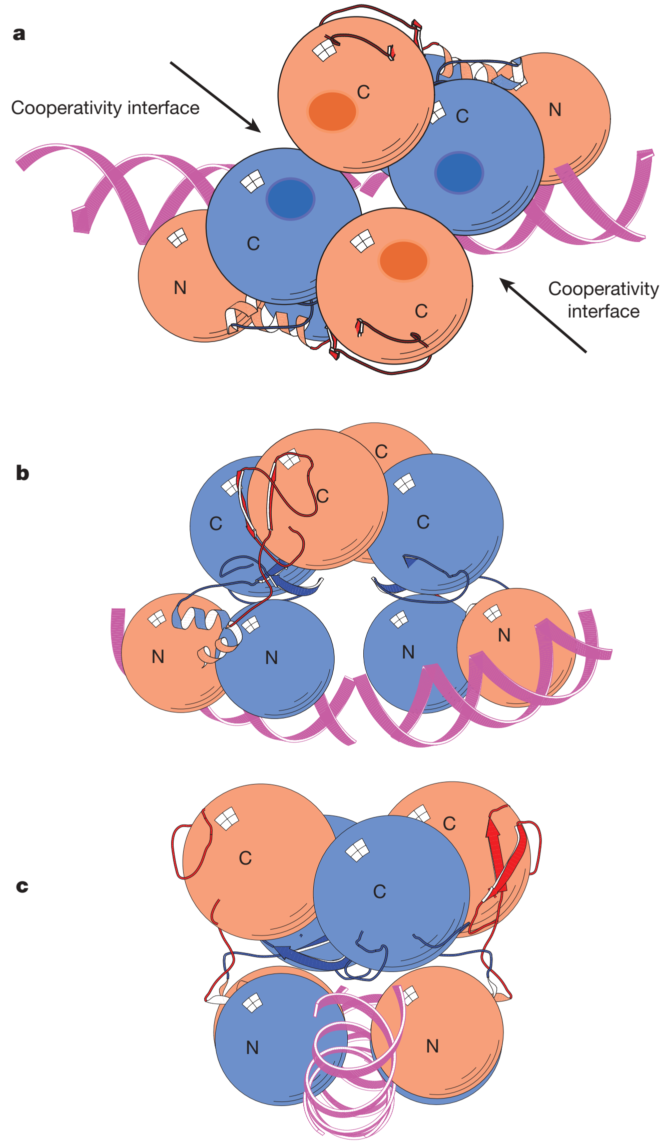



**Figure 3 | λ Repressor binds cooperatively to adjacent operator sites.** A model of the cooperativity complex was created by superimposing two intact dimers onto the structure of the CTD tetramer. **a**, Looking down on the cooperativity complex shows that the CTDs of the repressor dimers associate into a tetramer, as determined from previous crystal structures. The arrows draw attention to the cooperativity interface. Note that the amino-acid residues that mediate cooperativity comprise two discrete patches on the CTD surface (that is, four on the surface of each CTD dimer). When two dimers associate to form the cooperativity complex, only two of the four patches on each dimer become buried at the interface; the two remaining patches on each dimer (dark blue and dark pink) remain exposed to solvent, offering potential sites for further association (Fig. 4). **b**, The model of the cooperativity complex was rotated approximately 90° (about a horizontal axis) from the orientation shown in **a**. The model illustrates how two repressor dimers might bind to adjacent operator sites. The inter-operator spacing (centre-to-centre) in this model is about 22 bp, whereas the inter-operator spacing for adjacent operators in $O_R$ (and for $O_L1$ and $O_L2$) is either 23 or 24 bp. **c**, The cooperativity complex rotated 90° (about a vertical axis) from the orientation shown in **b** and viewed down the major axis of the DNA to illustrate that the NTDs bind laterally to the DNA, whereas the CTDs sit medially.

separating the NTD from the CTD. The connector (residues 93–131), which contains the cleavage site, consists of a small protease sensitive linker (residues 93–101) and the cleavage site region (CSR). Structurally, the CSR is an integral part of the CTD, forming a pair of antiparallel β-strands that drapes across its surface (see below). Cleavage occurs at a specific site (between Ala111 and Gly112) within a long loop (residues 106–126; the cleavage site loop) that connects the antiparallel β-strands of the CSR.

The most remarkable aspect of the intact dimeric structure is its lack of global symmetry. Unlike the cartoon representations, the structure reveals that the intact repressor does not have a single axis of symmetry. The two-fold axes that relate the NTD and CTD dimers are substantially skewed and are nearly orthogonal (Figs 1b–d and 2a, b). The interface between the two monomers is extensive, with about 2,200 $\text{Å}^2$ of buried surface area per monomer, of which less than one-half (about 900 $\text{Å}^2$) is due to the CTD–dimer interface. In this crystal structure, the NTD and CTD of one subunit of the dimer, but not the other, are in close contact, creating an interface that also buries about 900 $\text{Å}^2$ of surface area per monomer (blue subunit, Figs 1b–d and 2). For the most part the interactions between the NTD and the CTD are localized to the CSR, which provides the 'glue' that holds the NTD relative to the CTD (Fig. 2). In a space-filling representation of the dimer, this subunit of the repressor more closely resembles a cylinder than two discrete spheres (Fig. 2c). There are also interactions between the CSR of this subunit and the NTD of the dimer-related subunit (Fig. 2c), corresponding to about 350 $\text{Å}^2$ of buried surface area. As a consequence of this asymmetric dimer assembly, the two short linkers (residues 93–101) that connect the N-terminal and C-terminal domains have very different conformations. Although the asymmetric dimer assembly accounts for the pattern of cooperative repressor binding to adjacent operator sites (see below), we cannot exclude the possibility that crystal packing is responsible for this asymmetry. Additional experiments with mutant proteins might enable us to eliminate this possibility definitively.

The active site for the self-cleavage reaction is located in a shallow groove on the surface of the λ repressor CTD[7]. It comprises two essential residues (Ser149 and Lys192) that are strictly conserved among related self-cleaving repressors, including the bacterial repressor LexA[13]. In the case of LexA, a conformational equilibrium was shown to govern the rate of the self-cleavage reaction[14], and structures revealed the two conformations: one in which the cleavage loop folds over the active site (with the Ala–Gly bond appropriately positioned for cleavage) and one in which the scissile bond is about 20 Å from the catalytic centre[15]. The CSR in the structure of the intact λ repressor dimer adopts a conformation that resembles that of the LexA CSR in the non-cleavable conformation (Supplementary

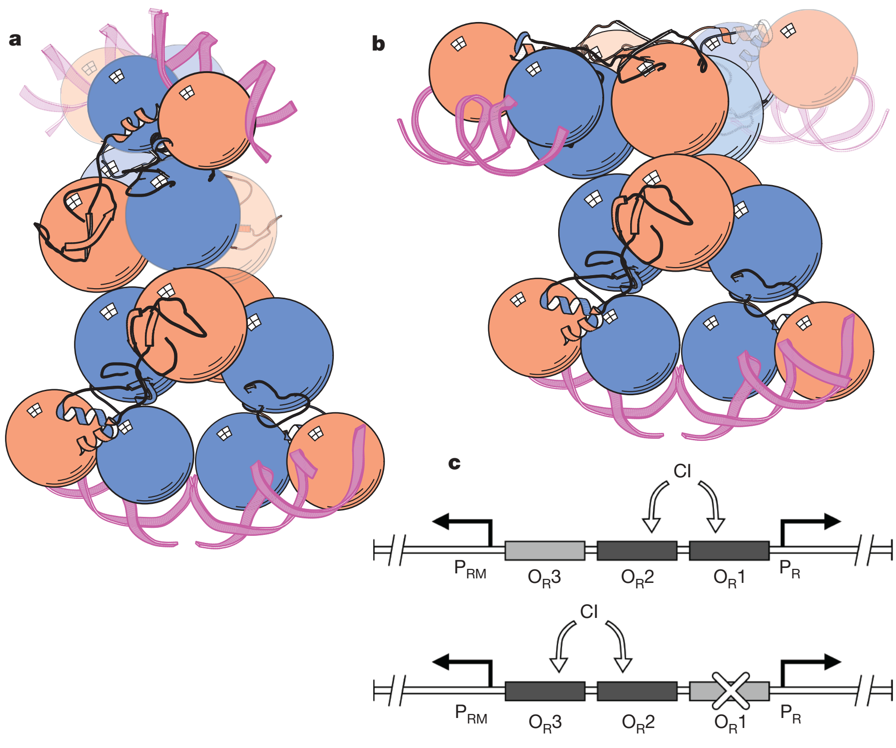



**Figure 4 | Models of the intact repressor octamer reveal the structural basis for alternate pairwise cooperativity.** **a**, Interaction of two DNA-bound tetramers. The model, which was created by superimposing two DNA-bound tetramers (see Fig. 3) on a structure of the λ repressor CTD octamer, illustrates why three interacting repressor dimers (shown in dark colour) cannot bind to three adjacent operators ($O_R1$, $O_R2$ and $O_R3$, for example). Each repressor dimer has two potential cooperativity interfaces[16], allowing it to interact with up to two additional dimers (consider, for example, the dimer at the bottom left, which is shown interacting with both the dimer at the bottom right and the darkly coloured dimer at the top). This superimposition of structures shows that only two of the three interacting dimers (those at the bottom) can be positioned at adjacent sites on the DNA. When the octamer is configured as shown, the third (top dark) and fourth (top pale) dimers are appropriately positioned to bind cooperatively to a distant pair of operator sites; the model thus shows how a pair of repressor dimers bound cooperatively to $O_R1$ and $O_R2$ can interact with a second pair bound cooperatively to $O_L1$ and $O_L2$, located 2.4 kb away. **b**, Alternative configuration for the intact repressor octamer. Because of the inherent asymmetry of the intact repressor dimer, two DNA-bound repressor dimers can be superimposed on the CTD tetramer in two different ways, resulting in two different configurations for the intact tetramer (one of which is depicted at the bottom, the other at the top). The superimposition depicted at the top does not allow for cooperative binding to adjacent operator sites; rather, the NTDs fan out in opposite directions. In this model, as well, however, the configuration of the three interacting dimers (shown in dark colour) is such that only two of the three (those at the bottom) can be positioned at adjacent sites on the DNA. **c**, Schematic of $O_R$ region that illustrates alternate pairwise cooperativity. A repressor (CI) dimer bound at $O_R2$ can interact with either another dimer bound at $O_R1$ or another dimer bound at $O_R3$, but not both simultaneously. On a wild-type $O_R$ region, sites $O_R1$ and $O_R2$ are occupied by a cooperatively bound pair of CI dimers (top), whereas when $O_R1$ is inactivated by mutation, sites $O_R2$ and $O_R3$ are occupied by a cooperatively bound pair of CI dimers (bottom)[4].

Fig. 2). Although it is likely that the mechanism by which cleavage is controlled is similar in these two systems, the higher-order interactions between repressor dimers push the equilibrium further towards the non-cleavable conformation in the case of λ repressor[12]. This is because the active site residues are buried within the cooperativity interface when the CTD dimers self-associate to form the cooperativity complex[7].

The unusual quaternary structure of the dimeric repressor provides insight into a long-standing problem of repressor function. The phage's $O_R$ region consists of three 17-bp operator sites ($O_R1$, $O_R2$ and $O_R3$), and repressor binds cooperatively to pairs of adjacent operators (normally to $O_R1$ and $O_R2$, but alternatively to $O_R2$ and $O_R3$ if $O_R1$ is inactivated by mutation)[4].

That is, the binding is 'alternate pairwise'; when the $O_R$ region is wild type, a dimer bound at $O_R2$ cannot interact simultaneously with dimers bound at $O_R1$ and $O_R3$. Based on the following modelling, we suggest that the inherent asymmetry of the DNA-bound repressor dimer (Figs 1 and 2) facilitates this pattern of alternate pairwise cooperativity, which is essential to the proper functioning of the genetic switch.

We first made a model of the pairwise cooperativity complex, which forms when the CTDs of adjacently bound repressor dimers self-associate and assemble as a tetramer (dimer of dimers)[7]. We modelled it by superimposing the CTDs of the intact dimeric repressor–operator complex onto the structure of the CTD tetramer[7]. This superimposition brings the DNA of one repressor–operator complex into rough alignment with that of the second repressor–operator complex, forming a nearly aligned continuous DNA helix (Fig. 3). Although the angular alignment of the operators in the model does not precisely match that of the natural operators (see Fig. 3 legend), this superimposition of structures provides the first plausible model for pairwise cooperativity. In particular, the model of the cooperativity complex suggests that when a pair of dimers in this conformation binds to adjacent operator sites, the CTDs are positioned medially, poised for tetramer formation (Fig. 3).

To address the issue of alternate pairwise cooperativity, we next took advantage of the structure of the CTD octamer[16] to model the placement of a third repressor–operator complex relative to the cooperatively bound dimer pair. That is, we superimposed the CTDs of the intact dimeric repressor–operator complex onto the structure of the CTD octamer[16] (Fig. 4). This superimposition can be done in two different ways, as depicted in Fig. 4a, b (see legend for details). Strikingly, in either case, the third repressor dimer (shown in dark colour at the top of each part of Fig. 4a, b) is positioned in such a way as to preclude its binding to an operator site located adjacent to the pair of operator sites shown at the bottom of each part of Fig. 4a, b. Thus, these models make it clear how the geometry of the CTD interactions precludes contact between a cooperatively bound dimer pair at $O_R1$ and $O_R2$ and an adjacently bound third dimer at $O_R3$. Rather, this superimposition of the CTDs of the intact dimeric repressor–operator complex onto the structure of the CTD octamer depicts the higher-order octameric complex that forms when the repressor dimer pair cooperatively bound at $O_R1$ and $O_R2$ interacts with a similar dimer pair cooperatively bound at $O_L1$ and $O_L2$, located 2.4 kilobases (kb) away (Fig. 4a)[5,17]. In conclusion, the current and previous structures thus account for both the short- and long-range repressor interactions that underlie the regulation of lysogeny.

## METHODS SUMMARY

The λ repressor protein with a single amino-acid substitution, D197G, was overproduced and purified. The repressor was crystallized with a 20-bp operator sequence using the hanging-drop vapour-diffusion method. Crystals grew in space group C222 with the following cell dimensions: $a = 74.46$ Å, $b = 160.97$ Å, $c = 134.83$ Å, $\alpha = \beta = \gamma = 90^\circ$ and diffracted to 4 Å. Data were collected at 100 K using a 345 MAR research image plate. The structure was solved by multiple isomorphous replacement (MIR) methods as implemented in SOLVE[18], and the phases improved by statistical density modification[19]. Previously determined structures of CTD (1F39) and NTD (1LMB) were fitted to the density using a genetic algorithm[20] and the model was initially refined using the rigid-body refinement with the Crystallography & NMR System[21]. Crystallographic and refinement statistics are in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 29 July 2007; accepted 13 February 2008.**


1. Ptashne, M. *A Genetic Switch* (Cold Spring Harbor Press, Cold Spring Harbor, New York, 2004).
2. Maniatis, T. & Ptashne, M. Multiple repressor binding at the operators in bacteriophage lambda. *Proc. Natl Acad. Sci. USA* **70,** 1531–1535 (1973).
3. Maniatis, T. *et al.* Recognition sequences of repressor and polymerase in the operators of bacteriophage lambda. *Cell* **5,** 109–113 (1975).
4. Johnson, A. D., Meyer, B. J. & Ptashne, M. Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc. Natl Acad. Sci. USA* **76,** 5061–5065 (1979).
5. Dodd, I. B., Perkins, A. J., Tsemitsidis, D. & Egan, J. B. Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev.* **15,** 3013–3022 (2001).
6. Pabo, C. O. & Lewis, M. The operator-binding domain of λ repressor: structure and DNA recognition. *Nature* **298,** 443–447 (1982).
7. Bell, C. E., Frescura, P., Hochschild, A. & Lewis, M. Crystal structure of the lambda repressor C-terminal domain provides a model for cooperative operator binding. *Cell* **101,** 801–811 (2000).
8. Whipple, F. W., Kuldell, N. H., Cheatham, L. A. & Hochschild, A. Specificity determinants for the interaction of lambda repressor and P22 repressor dimers. *Genes Dev.* **8,** 1212–1223 (1994).
9. Pabo, C. O., Sauer, R. T., Sturtevant, J. M. & Ptashne, M. The lambda repressor contains two domains. *Proc. Natl Acad. Sci. USA* **76,** 1608–1612 (1979).
10. Jordan, S. R. & Pabo, C. O. Structure of the lambda complex at 2.5 Å resolution: details of the repressor–operator interactions. *Science* **242,** 893–899 (1988).
11. Sauer, R. T., Ross, M. J. & Ptashne, M. Cleavage of the lambda and P22 repressors by recA protein. *J. Biol. Chem.* **257,** 4458–4462 (1982).
12. Little, J. W. Autodigestion of lexA and phage lambda repressors. *Proc. Natl Acad. Sci. USA* **81,** 1375–1379 (1984).
13. Slilaty, S. N. & Little, J. W. Lysine-156 and serine-119 are required for LexA repressor cleavage: a possible mechanism. *Proc. Natl Acad. Sci. USA* **84,** 3987–3991 (1987).
14. Roland, K. L., Smith, M. H., Rupley, J. A. & Little, J. W. In vitro analysis of mutant LexA proteins with an increased rate of specific cleavage. *J. Mol. Biol.* **228,** 395–408 (1992).
15. Luo, Y. *et al.* Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell* **106,** 585–594 (2001).
16. Bell, C. E. & Lewis, M. Crystal structure of the lambda repressor C-terminal domain octamer. *J. Mol. Biol.* **314,** 1127–1136 (2001).
17. Revet, B., von Wilcken-Bergmann, B., Bessert, H., Barker, A. & Muller-Hill, B. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Curr. Biol.* **9,** 151–154 (1999).
18. Terwiliger, T. C. & Berendzen, J. Automated structure solution for MIR and NAD. *Acta Crystallogr. D* **55,** 849–861 (1999).
19. Winn, M. D. An overview of the CCP4 project in protein crystallography: an example of a collaborative project. *J. Synchrotron Radiat.* **10,** 23–25 (2003).
20. Chang, G. & Lewis, M. Molecular replacement using genetic algorithms. *Acta Crystallogr. D* **53,** 279–289 (1997).
21. Brunger, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank S. Garrity for reading this manuscript.

**Author Information** Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under the accession number 3BDN. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.L. (lewis@mail.med.upenn.edu).

## METHODS

Codons 1–236 of the λ repressor with a single amino-acid substitution, D197G, were cloned into a derivative of pET 17b (Novagen) to produce a gene that codes for a repressor with a non-cleavable $His_6$ tag at the C terminus. The protein was produced in *E. coli* BL21 and purified using Ni-NTA affinity column (Qiagen). The purified λ repressor D197G was dialysed overnight against buffer that contained 200 mM Tris, 200 mM KCl, 50 mM EDTA, 3 mM DTT (pH 7.4) and concentrated to 80 mg $ml^{-1}$. The sequence of the DNA used in crystallization, supplied by the Nucleic Acid Facility, Department of Chemistry, University of Pennsylvania, contained 20 bases with a total of 19 complementary base pairs with one base overhanging from each strand. Centred on the DNA is the 17-bp $O_L1$ sequence, which is the operator that has the highest binding affinity for λ repressor:

5′ TATATCACCGCCAGTGGTAT 3′
3′ TATAGTGGCGGTCACCATAA 5′

To ensure the integrity of the protein and the DNA used for crystallization, the binding of repressor to the operator was confirmed by running a native agarose gel (1.4%). The protein–DNA complex was prepared each time before crystallization. The complex was prepared with the λ repressor D197G dimer and the purified operator in a 1:1.2 molar ratio by diluting the repressor to 10 mg $ml^{-1}$ before mixing. The crystals were obtained by hanging-drop vapour diffusion, using 2 μl of the complex solution and 2 μl of the well solution. Crystallization trays were kept undisturbed in an incubator at 21 °C. The crystallizing condition was 14% PEG 4000, 0.1 M sodium acetate (pH 4.6), 0.2 M ammonium sulphate, 0.05 M magnesium chloride, 20% (v/v) glycerol and 1 mM spermine. Crystals, which usually take two to three days to form and about a week to mature, grow in space group *C*222 with the following cell dimensions: $a = 74.458$ Å, $b = 160.969$ Å, $c = 134.83$ Å, $\alpha = \beta = \gamma = 90°$ and diffract to 4 Å both in house and at high-energy X-ray sources. Native data were collected at 100 K using a 345 MAR research image plate mounted on a Rigaku RU-3HR rotating anode generator (University of Pennsylvania). Two hundred frames were collected at 1° oscillation, 30-min exposure per frame, at 400 mm from the detector. A native data set was collected to 3.8 Å at 100 K at the National Synchrotron Light Source beamline X25 using a charge-coupled device detector. All data were processed and scaled using DENZO and SCALEPACK[22].

The structure was solved by MIR. The screening for heavy atoms was greatly simplified by using derivatives that were successful for solving the isolated NTD[6] and CTD[7]. The NTD was phased with $K_2PtCl_4$ and a multiple anomalous displacement data set of the CTD was collected from a $Hg(OAc)_2$-derivatized crystal. To prepare the derivatives, stock solutions of 20 mM of each heavy atom were created. One microlitre of heavy-atom stock solution was mixed with 9 μl drawn from the mother liquor of a mature crystal to be derivatized to make a 2 mM solution. An appropriate amount was then added, slowly and carefully by a pipette, to the drop containing mature crystals to make the final concentration of 0.5 or 1 mM in the drop. The coverslip was then returned to the well.

Both the mercury and platinum derivative data sets were collected at 100 K to 4.5 Å using a 345 MAR research image plate mounted on a Rigaku RU-3HR rotating anode generator (University of Pennsylvania). Two hundred frames were collected at 1° oscillation, 15-min exposure per frame, at 400 mm from the detector.

The positions of the heavy atoms were determined and preliminary phases generated using SOLVE[18]. The validity of the sites was checked by cross Fourier calculations, and the MIR phases were calculated using the CCP4 suite[19]; the phasing power was 1.33 for mercury and 1.21 for platinum. The initial phases, improved by statistical density modification, were used to generate an initial electron-density map to 4.5 Å. As anticipated, the mercury heavy atoms bound to the CTD of the repressor as they did in the original structure. The platinum derivative, however, did not occupy the same position in the intact repressor as it did in the NTD structure. The CTD and NTD were readily recognizable (Supplementary Fig. 1).

The previously solved structures of CTD (1F39) and NTD (1LMB) were fitted to the density using a genetic algorithm we developed specifically for this purpose[20]. The heavy atom sites were then confirmed by calculating a difference Fourier transform using phases calculated from the model. The preliminary model of the repressor was used to determine the non-crystallographic symmetry of the repressor; the initial phases were further improved by using the model as a mask and averaging the density of the NTD and the CTD[19]. The electron density allowed us to trace the polypeptide chain of the CSR using the graphics program COOT[23]. Supplementary Fig. 1 illustrates the fit of the cleavage-site region loop to the electron density.

The structure was initially refined using the rigid body refinement with Crystallography & NMR System[21], where the NTDs, CTDs and DNA were treated as separate rigid bodies. This reduced the *R* factor and free *R* factor significantly ($R_{free}$ 45–39%; *R* 44–38%). At this point the model for the repressor was examined using difference Fourier maps. There were loops in the CTD that clearly were displaced; these were corrected. In subsequent rounds of refinement at 3.9 Å, we used simulated annealing and tight NCS restraints as well as targeted restraints on the DNA, which reduced $R_{free}$ to 35% and *R* to 29%. New $2F_o - F_c$ electron-density maps showed clear density for the main-chain atoms but the positions of the side-chain atoms were not readily apparent. At this resolution the refinement must be heavily restrained as the number of parameters to observations is greater than 2; we chose not to push the refinement any further. The root mean squared deviation of the backbone atoms of the NTD and CTD dimers is less than 0.2 Å, and the stereochemistry of the model is consistent with known structure.

22. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
23. Emsley, P. & Cowtan, K. COOT: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).

# naturejobs

## JOBS OF THE WEEK

China's growing research and development (R&D) enterprise is something of a paradox. The country is experiencing incredible economic growth, largely the result of a manufacturing boom, and the government continues to increase R&D funding. But the country's students and young scientists are struggling to cling to their dreams of working independently in academia or pursuing their preferred science professions. Meanwhile, foreign-trained PhD students are often still favoured over home-grown talent (see page 1028).

Major reforms to China's higher-education system in the late 1990s improved its capacity to teach students in a variety of disciplines, and enrolment has risen sharply — up from 0.3 million to 13.3 million for undergraduates between 1998 and 2004. A similar trend can be seen in graduate schools, but graduate studies, including those in science and technology, present a stumbling block for career-seekers in the form of poor pay.

Many of China's top students still head overseas, where the United States is the most popular destination. A report from the US Council of Graduate Schools, released on 14 April, suggests that although graduate-school applications from China have slowed somewhat, they are still significant — particularly when compared with applications from other nations.

Other countries also attract plenty of Chinese talent. Around 19% of Canada's science and engineering graduate students were from China in 2004. And of the more than 18,000 foreign graduate students enrolled in Japanese universities in 2005, nearly 10,000 of them were from China.

Perhaps it is only a matter of time before the Chinese government raises postdoc salaries and graduate-student stipends. And maybe there is no rush to stem an exodus of students who receive good training and cross-cultural experiences abroad — as long, that is, as China finds a way to create jobs to accommodate the increasing number who are returning.

**Gene Russo is editor of *Naturejobs*.**

CONTACTS
**Editor:** Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

**European Sales Manager:**
Andy Douglas (4975)
e-mail: a.douglas@nature.com
**Business Development Manager:**
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
**France/Switzerland/Belgium:**
Muriel Lestringuez (4994)
**Southwest UK/RoW:** Nils Moeller (4953)
**Scandinavia/Spain/Portugal/Italy:**
Evelina Rubio-Hakansson (4973)
**Northeast UK/Ireland:**
Matthew Ward (+44 (0) 20 7014 4059)
**North Germany/The Netherlands:**
Reya Silao (4970)
**South Germany/Austria:**
Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:** Tom Hancock
***Naturejobs* online production:** Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

**US Sales Manager:** Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
**Business Development Manager, Greater China/Singapore:**
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

NATURE|Vol 452|24 April 2008

# CHINA'S CHALLENGE

**Young researchers in China face stiff competition as they strive to establish labs or find other scientific careers. Wei Zeng explores what it takes to succeed as a Chinese scientist.**

Scientific enterprise is undergoing a much-touted expansion in China, fuelled by astronomical economic growth and a seemingly unlimited supply of talent. But graduate students and postdocs in science and engineering face numerous hurdles. For many, their dreams of becoming an academic researcher seem out of reach, owing to problems with funding, training and bureaucracy. "If you are really interested in academia, it is worth trying. Otherwise, run away," says Yi Wang, a third-year biology graduate student at Peking University in Beijing.

The Chinese government continues to invest heavily in science and engineering. In 2006, it increased spending on research and development (R&D) to 1.42% of gross domestic product (GDP). According to China's National Bureau of Statistics, this is a historical high. Although the percentage is lower than that of the United States or Japan, the total expenditure of around 300 billion renminbi (US$43 billion), is second only to the United States. More than 700 multinational companies have built R&D centres in China.

But limited government support and an excess of young researchers have made prospects dim for many. The job market has become more crowded as Chinese researchers return from overseas. And although the government has expanded PhD and postdoc programmes in the past few years, there's still a shortage of full-time research vacancies.

Postdoctoral positions were instituted in China in 1985, largely on the suggestion of Chinese-born Nobel physics laureate Tsung Dao Lee. By the beginning of 2007, China had put a total of 45,000 postdocs through the system (see 'The rise of the postdoc'). In 2006, the latest year for which official statistics are available, there were more than twice as many postdocs who were in their posts or looking for jobs than in 2002. Some 40% were in engineering, 20% in science and 10% in medicine, according to the 2006 figures.

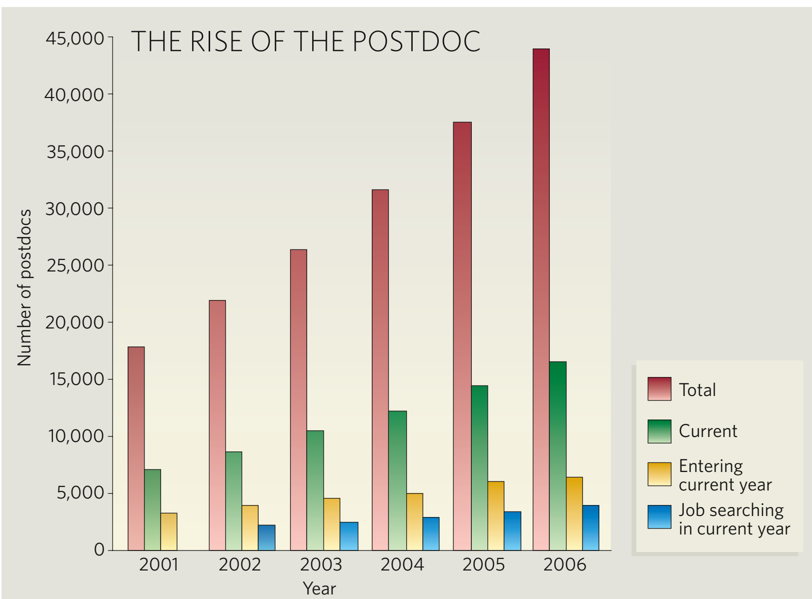
Cong Cao, a research fellow at the State University of New York, who studies workforce-related issues in China, says that roughly 31% of China's science and technology postdocs work in basic science and the rest in applied science. Although the percentage going to industry, government, the military or a second postdoc changed little between 2003 and 2006, the percentage going to academia dropped, and the percentage of postdocs taking other positions increased (see 'Where they go'). Either unconventional career paths are on the upswing, or the market for conventional positions in academia and overseas has become saturated — or both.

## Prodigal sons

The top talent still heads abroad. Only 11% of China's PhDs continue their research in Chinese institutions, according to the Ministry of Science and Technology. "Postdoctoral positions in China are not as attractive as overseas ones in terms of both salary and scientific training if the PhDs want to go further in their scientific career," says Gang Pei, president of the Shanghai Institutes for Biological Science (SIBS) and president-elect of Tongji University in Shanghai.

To some extent, the Chinese government encourages the flight overseas. Government initiatives such as the '100 Talents', a Chinese Academy of Sciences programme funding 100 scholars per year, typically favour those with foreign research experience, says Cao. Neuroscientist Yi Rao, who moved back to Beijing in 2007 after 24 years in the United States, acknowledges this stigma against young scientists who have not studied or worked abroad, but suggests that things are changing. Rao is dean of Peking University's life-sciences department and deputy director of the National Institute of Biological Sciences (NIBS), a government-financed institute that opened in 2003.

More of a problem, Rao laments, is the lack of an intellectual lab culture in many Chinese institutions. One of the first things he did upon arrival at Peking University was to start a regular seminar series. Most labs in China




SOURCE: CHINA POSTDOCTOR ORGANIZATION

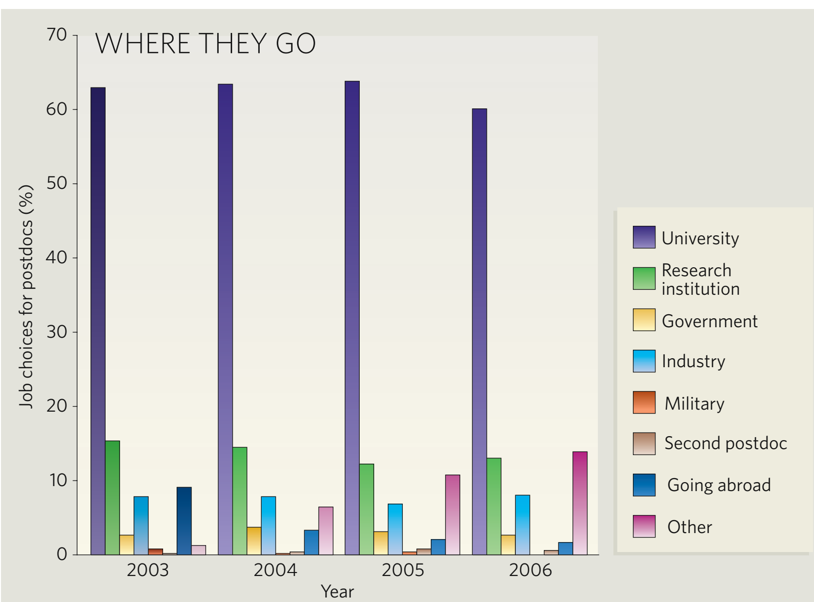

Science graduate students at Peking University are seeking sound career paths.


LIPING WEI/PEKING UNIV.

are focused on producing papers and results, rather than interacting with other labs, Rao says. "If investigators treat science as hard labour, they do not enjoy science," he says. "And a natural consequence is that their students will soon see through their motivation and lose interest."

Notoriously low stipends exacerbate the difficulty of staying in academia. The typical PhD student makes between 500 and 1,200 renminbi a month — barely enough to survive on in big cities such as Beijing. More than 50 biologists and mathematicians recently wrote to the government requesting increased support for graduate students. The ministries are discussing initiatives, but no definite steps have been taken, according to Rao, who helped lead the appeal. According to China's *Xinmin* newspaper, Yuliang Yang, head of the office of academic degrees committee of the state council, told a conference in November 2007 that China's PhD students cannot live a decent life normal for people in their 20s and 30s. At a December 2007 workshop on graduate education at Peking University, administrators

**"If investigators treat science as hard labour, their students will soon see through their motivation and lose interest."** —Yi Rao




SOURCE: CHINA POSTDOCTOR ORGANIZATION

at top Chinese universities expressed similar concerns. Taking a second job is rarely an option for dedicated graduate students, notes Cao, as they are expected to publish two or three papers within three years. Young researchers do not fare much better once they get their degrees: typical salaries for entry-level PhDs and postdocs are 2,000–5,000 renminbi a month, with a rare few making thousands more at high-tech companies.

## Supply side

The government controls the supply of postdoc positions, allotting a certain number to each institution and often limiting their duration to two to three years — a small time window to finish a complex research project intended to propel the researcher into an independent career. This isn't a problem everywhere, notes Rao — NIBS for example has no time limit. Postdoc numbers will continue to increase — the Chinese Academy of Sciences runs more than 100 postdoctoral programmes, which accommodate 20% of all Chinese postdocs. Its total budget for postdoc support has increased from 1 million renminbi in 1985 to 3 million today and it plans to support some 4,000 postdocs by 2010, three times its current number.

But many still struggle. According to Pei, principal investigators at SIBS spend their own grant money to support 70–80 postdocs a year and raise the average salary. They are also trying to extend the duration of a postdoc in an effort to make their PhDs and postdocs more competitive in the job market.

Weiwei Liu, one of Rao's graduate students, plans to apply for overseas postdoc positions, but worries about her job prospects. "As far as I know, the opportunities are limited and the competition is great," she says. She sees limited opportunities in China in academia and an underdeveloped biomedical industry.

Still, companies could be part of the solution. A few do fund university postdocs to address specific research problems. Others earn certification to start their own postdoc positions. Although this leaves companies setting the agenda, Yi Bing Duan, an assistant professor at the Institute of Policy and Management at the Chinese Academy of Sciences, calls it a potential "triple win" — universities get money, postdocs get support and companies get their research questions addressed. Duan would like to see postdocs get their own organization, something like the National Postdoctoral Association in the United States, which would identify postdoc gripes and work with the government to find solutions.

Despite the frustrations, researchers with dreams of pursing basic science do persevere. Meng Gao, a PhD student in bioinformatics at Peking University, is cautiously optimistic. She pictures herself setting up her own lab and spending 20 years on one project. "The employment situation is especially tough for women scientists," she says. "I know many of them are facing difficulties and are planning to give up, or even have left."

But there are success stories. Jiang Zhang was recently appointed an assistant mathematics professor at Renmin University in Beijing after a two-year postdoc at the Chinese Academy of Sciences. He enjoys the freedom of academia. "I decide when to do what project and how to do it," he says. "The money is not good but is acceptable. And I believe not only the money but the overall situation in academia will get better. Even if one day I go to industry, I still want to do research." ■

**Wei Zeng is a freelance writer for *Nature China*.**

NATURE|Vol 452|24 April 2008

# MOVERS

**Marja Makarow, chief executive, European Science Foundation, Strasbourg, France**

**2003-07** Vice-rector, University of Helsinki
**2003-** (on leave of absence until 2012) Professor of applied biochemistry and molecular biology, University of Helsinki
**2000-04** Research director, Program in Cellular Biotechnology, University of Helsinki

Marja Makarow's early interest in linguistics soon gave way to an interest in the sciences after she fell in love with the word 'biochemistry'. But her ability to speak several languages and understand multiple fields will serve the Finnish biochemist well in her new role as chief executive of the European Science Foundation (ESF).

Makarow is charged with finding new ways to encourage cooperation across borders (both political and scientific) to support pan-European and interdisciplinary programmes. The ESF is not a funding agency, although its 77 member bodies include major funding agencies and research institutions from 30 countries. Rather, it facilitates networking, runs conferences and publishes position papers.

The University of Copenhagen's Liselotte Højgaard, who was on the search committee that selected Makarow, cites her strong academic background, international experience and leadership roles. She recalls seeing Makarow present strategies in a new white paper from the ESF's European Medical Research Councils. "She was very good at facilitating it, in a friendly, kind, polite and still firm way," says Højgaard. "She doesn't come across as a bulldozer. She listens."

Makarow studied biochemistry at the University of Helsinki, where she did postdoctoral work in virology, and went on to the European Molecular Biology Laboratory (EMBL) in Heidelberg. She was a group leader at the University of Helsinki's Institute of Biotechnology before becoming the institute's research director, an ascent interrupted only by a stint at the University of Kuopio.

Ten years ago, Makarow ventured into policy, serving on the Academy of Finland's Research Council for Health, the main funding agency for biomedical research. Next, she served on the governing body of EMBL and its multinational funding organization, a window on the European grant funding system. As vice-rector at the University of Helsinki, she oversaw research in many scientific disciplines. "When you give the little finger, the hand often goes," says Makarow, recounting the many positions she has held in addition to running her own lab.

Højgaard uses the word *sisu* to describe Makarow. "It means the ability to go through stone," she says, noting that the word often refers to a personality trait that Finns have developed during centuries of surviving cold, dark winters and holding Russian incursions at bay. "It's a tough job to be the chief executive of the ESF, but she can do it. She really has *sisu*." ■

**Jill U. Adams**

# MENTORS & PROTÉGÉS

## A guiding COMPASS

As I contemplated pursuing a PhD in cell biology, I looked back at my own undergraduate experience and decided that scientists in minority groups needed a more supportive pathway. With the help of associate deans and advisers at Carnegie Mellon University, I set up COMPASS (Coaching Minority Progress and Academic Success in Science), a mentoring programme led by older students that helps incoming minority students. That programme now has 15 members. Its genesis was rooted in my own struggles adjusting to a rigorous academic atmosphere.

During my first year as an undergraduate, I often felt overwhelmed and unsure about my abilities as a student. I was lucky to meet Malika Jeffries-EL, an African-American postdoc in chemistry, whose encouragement convinced me that my hard work would pay off. Some struggling, she said, was natural. She and I had much in common. She too had been the only black student in a science class. She knew the pressure to avoid failure. Communicating with a woman of colour who had earned an advanced science degree helped me see a path to success.

My work improved in my second year. My research project, analysing the interactions of proteins involved in ribosome assembly, was an eye-opening experience. I enjoyed the various biochemical and genetic techniques used to study cellular mechanisms; I was stimulated by the questions that arose from analysing data; and I became interested in solving biological problems and addressing deeper questions at the PhD level. This year I was fortunate to be awarded a Howard Hughes Medical Institute Gilliam Fellowship for Advanced Study, which will fund the entirety of my graduate-school experience at the school of my choosing. I hope one day to become a cell-biology professor.

My experience of mentoring inspired me to start COMPASS, which enables incoming minority students to network with current students, key advisers and faculty members. Students are contacted the summer before their first year and invited to participate. The programme helps me share my insights in a supportive network.

This autumn, I plan to study regulation of the cell cycle at Stanford University. I will continue to serve as a mentor and adviser to COMPASS, hoping to help young minority scientists develop their own paths to success. ■

**Betty Mbom is a senior at Carnegie Mellon University.**

**POSTDOC JOURNAL**

## In the name of science

This month I ran more than 20 metres carrying fresh gelada faeces on a rock (trying to photograph the defecator) and learned to tell females apart by the length of their nipples. It's all in the name of science. I love my job! I am living in a monkey soap opera — gelada baboons quarrel and have sex daily, providing us scientific voyeurs with hours of data-gathering.

All I miss is free time. I'd forgotten how much time is consumed by the process of gathering and digitizing behavioural data. Still, I'm an absolute nerd. I find it exhilarating to see the information increase, and to transcribe gelada vocalizations painstakingly recorded through the hazards of windy weather and tourist encounters. New ideas are already brewing in my mind, new things on which to focus, new questions to ask. (Do rank, spacing and/or hormones affect their contact calling?) My brain works in spite of the oxygen deficit!

Perhaps I am sacrificing some ordinary social skills, though. I am learning to identify some 100 geladas by their unique individual quirks or characteristics. The downside? I seem to be classifying new people as if they were geladas. Just the other day I caught myself comparing one man's straggly whiskers with another's skewed moustache and thinking, 'These males are easy to tell apart!' I foresee problems readjusting to the human soap opera when I leave the field. ■

**Aliza le Roux is a postdoctoral fellow in animal behaviour at the University of Michigan.**

NATURE|Vol 452|24 April 2008

# SETI for profit

A capital idea.

**Gregory Benford**

Multi-billionaire Clyde Kraft's announcement that his team had detected a SETI signal shocked the world. He claimed to have found a plainly artificial broadband signal from a star well over a thousand light years away.

His astronomers ranked behind him confirmed the discovery. They gave pulse spectra, but without labels that would let anyone find or decipher it. They omitted the celestial location or frequency. Then Kraft stepped back to the microphone and gave his terms.

"The real news is that we've found a message in the megahertz-wide signal. It's intermittently on, hard to find. But we held steady, studied the galactic plane for over a year." Kraft paused for effect, giving the world a thin smile. "Then — there it was! In less than a week my fine team deciphered it. Not hard, really. It's amazing. And … it's a warning."

Consternation. Shouts from the press, who thought they were attending a business briefing. Kraft wouldn't reveal the message or how to find the source. He scowled. "I did this as an … investment. Pay me a finder's fee, say half a billion dollars, and we'll talk." With another smile he left the dais and vanished from public view.

Around the world scientists rebuked him. "To make such a discovery and keep it secret violates the code of science," the head of the National Academy pronounced. "And to expect profit!"

Parliaments voted to condemn Kraft. "The most important discovery in history cannot be private property," declared the Pope.

Most news shows carried the story as a scandal, focusing on Kraft's known cut-throat ways of gaining advantage. He had used head-to-head methods in stochastic arbitrage trading, making him Forbes 2013 'King Competitor'. Even the *Wall Street Journal* went tsk-tsk. *The Times* in London had a headline three inches high: "What is the warning?"

Kraft wouldn't say. Appeals by politicians, wise men and philosophers got no answer at all.

Worldwide, astronomers condemned Kraft — and got to work. They pored over older radio-telescope data, looking for sources 1 megahertz wide that came and went. There were several and they earned front-page coverage.

JACEY

Dishes trained on these candidates listened constantly, to the limits of their sensitivity. Kraft would not reveal the discovery frequency so they scanned through their entire range. This was a marked change from conventional SETI, which had gathered mostly at the 'water hole' between 1 and 2 gigahertz.

Observers with telescope time donated it to the search for the mysterious source. Amateurs erected their own 3-metre radio dishes in backyards and got in on the hunt with gear they bought over the Internet.

Kraft's in-house astronomers were sound scientists who had lost out in the academic horse race. They refused even to discuss the discovery. SETI had been losing momentum since 2010, the 50th anniversary of its first search. Kraft had sought to reinvigorate the field, they said, and did so with what was to him pocket change, only $100 million.

Scientific professional societies turned on these astronomers. Furious executives in the AAAS, the IAU and other bodies revoked the astronomers' memberships. Mobs rioted outside their homes. 'Free ET!' shouted a popular sign the crowds carried in capitals around the world.

Kraft stayed in hiding. Pressure built. The Kraft astronomers also vanished, harried by incessant demands.

Months passed. The older observed signals yielded nothing. The United Nations debated whether it could or should confiscate Kraft's holdings within its reach — only to find that he had moved them into secure accounts. Nations began to negotiate how to share out the cost of meeting Kraft's price.

Kraft dodged negotiations. Rumours spread that he had become a greybearded hermit, like Howard Hughes. Reports located him in Switzerland and North Korea.

Then the Australian Parkes radio telescope picked up a quick, millisecond signal not far from the line of sight towards the galactic centre. It was 3 megahertz wide, not 1, but it did have a signal.

Parkes released the pattern immediately. Thousands of scientists leapt to decode the message. Within days they had a coherent form, mostly details about a solar system similar to ours. It identified a race of elephant-like creatures complete with trunks, legs and two grasping hands at the shoulders. Their society was very old and might be extinct. There were hints that robots tended the message-sending beacon.

But there was no warning in the message at all.

With 30 minutes' notice, Kraft appeared at a press conference in Singapore. He admitted that there was no warning, the bandwidth was different — and in fact he had never detected a signal at all.

"I wanted to goose the world," he said. "Get 'em started. SETI needed that a lot more than it needed more money. Now we have something and it's time to get to work. I'll stand aside."

He left the room in an uproar. But that was nothing when, three years later, he won a Nobel, along with Parkes Observatory. ■

**Gregory Benford is professor of physics and astronomy at the University of California, Irvine, and a novelist.**

FUTURES